# ASSISTED VIEWPOINT INTERACTION FOR 3D VISUALIZATIONS

by

Stephen Benjamin Hughes

B.S. Hons. in Computer Science, Dickinson College, 1995

M.S. in Computer Science, Indiana University, Bloomington, 1997

Submitted to the Graduate Faculty of

The School of Information Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2005

UNIVERSITY OF PITTSBURGH

SCHOOL OF INFORMATION SCIENCES

This dissertation was presented

by

Stephen Benjamin Hughes

it was defended on

5 August, 2005

and approved by


Dr. Michael Lewis
Department of Information Science and Telecommunications
Dissertation Advisor

Dr. Peter Brusilovsky
Department of Information Science and Telecommunications

Dr. Shi-Kuo Chang
Department of Computer Science

Dr. Stephen C. Hirtle
Department of Information Science and Telecommunications

Dr. Michael B. Spring
Department of Information Science and Telecommunications

ASSISTED VIEWPOINT INTERACTION FOR 3D VISUALIZATIONS

Stephen B. Hughes

University of Pittsburgh, 2005

Many three-dimensional visualizations are characterized by the use of a mobile viewpoint that offers multiple perspectives on a set of visual information. To effectively control the viewpoint, the viewer must simultaneously manage the cognitive tasks of understanding the layout of the environment, and knowing where to look to find relevant information, along with mastering the physical interaction required to position the viewpoint in meaningful locations. Numerous systems attempt to address these problems by catering to two extremes: simplified controls or direct presentation. This research attempts to promote hybrid interfaces that offer a supportive, yet unscripted exploration of a virtual environment.

Attentive navigation is a specific technique designed to actively redirect viewers' attention while accommodating their independence. User-evaluation shows that this technique effectively facilitates several visualization tasks including landmark recognition, survey knowledge acquisition, and search sensitivity. Unfortunately, it also proves to be excessively intrusive, leading viewers to occasionally struggle for control of the viewpoint. Additional design iterations suggest that formalized coordination protocols between the viewer and the automation can mute the shortcomings and enhance the effectiveness of the initial attentive navigation design.

The implications of this research generalize to inform the broader requirements for Human-Automation interaction through the visual channel. Potential applications span a number of fields, including visual representations of abstract information, 3D modeling, virtual environments, and teleoperation experiences.

**ACKNOWLEDGEMENTS**

Standing on the shoulders of giants, I would like to express my gratitude to all who have provided the support to complete this dissertation.

Throughout my doctoral studies, the members of my committee have stimulated and guided my thoughts. I would especially like to thank Dr. Michael Lewis for his years of commitment and collaboration. From the beginning, he has provided an environment that fostered intellectual growth through a mix of freedom to explore new ideas, practical research experiences, and opportunities to join the larger community of academics. I appreciate Dr. Michael Spring for his engaging conversations and broad vision of Information Science. Dr. Peter Brusilovsky shifted my thinking about recommendation systems and personalization. I value Dr. Stephen Hirtle's background in spatial cognition and human information processing. Dr. Shi-Kuo Chang shares an interest in visual information and communication; his classes were both enjoyable and enlightening. I am proud to be associated with this group of professors and to be a product of their combined intellects.

Doctoral studies certainly shape an individual, but at a more foundational level, I would also like to thank all the people who forged my intellectual curiosity. This includes almost all of my teachers from elementary school to the present, but some distinctly stand out for the impact that they have made. Warren Nelson believed that teaching ninth graders symbolic logic in English class would enhance their writing skills. His classes brought about what can best be described as a "rational awakening", for which I am eternally grateful. David Blakney opened the door to structured thinking and problem solving through high school programming classes. He offered flexible learning opportunities and independent studies that I believe were instrumental in my current achievement. At Dickinson College, Dr. Anil Shende introduced me to academic life. I am thankful to him for including me in his research and getting me started on the path to professional scholarship. During my Master's program, Dr. Andrew Hanson exposed me to the potential of visualization research and pioneered the work that led to this study. I am indebted to him and his collaborator Dr. Eric Wernert for allowing me to participate in their pursuit of knowledge.

When some people come into your life, it is evident that you are better for having known them. I consider myself fortunate to recognize some of these individuals.

For the greater part of my life, Demetri Karousos has been engaging my mind, challenging my understanding of the world and honing my competitive spirit. The word to describe our relationship eludes me – it is certainly stronger than friendship. Regardless, I hold it and him in the highest esteem.

When I was in 10th grade, Dr. William J. Sacco hired me to produce business charts and assist with developing glyphs to understand hospital trauma data. That experience was so much more than an after-school job; it gave me an invaluable opportunity to learn. I thank him for the confidence that he placed in me and for the inspiration set by his example.

For the past six years, I have known Dr. Anthony Debons as the ever-present emeritus professor who inevitably has a question for the speaker. I thank him for his philosophy and for offering a shining model of academic life.

I treasure the conversation, speculation and stimulation from my fellow graduate students: Maria Harrington, Cholyeun Hongpaisanwiwat, Jeffrey Jacobson, Denis Nkweteyim and Jijun Wang. The future is brighter, knowing that peers like this gather on the horizon.

Anyone who has completed a dissertation knows that a critical ingredient is dedication and support from family. In that regard, I think that I am fortunate beyond compare. My parents, Charles and Susan Hughes and parents-in-law, Garold and Linda Koopman have been nothing but encouraging and patient in this and all other endeavors that I have chosen to pursue.

Finally, and most importantly I want to thank my wife, Elizabeth. Throughout everything, I know that she is one constant that will always be there for me. I truly believe that I could not have done this without her by my side – nor would I have wanted to. I love you forever.

**To Ethan Edwin Hughes**

*Me love you BIG much too!*

# 1. Introduction

"The purpose of computing is insight, not numbers".

~Roger Hamming [58]

Students of visualization will invariably encounter this quote – usually after reading less than a half-dozen papers. This is particularly interesting since this banner statement does not directly refer to the field that has appropriated it. However, visualizations offer such an intuitive appeal as a method for providing insight, that nobody questions the suitability of this citation. This intuition is so strong that people also readily credit the saying, "a picture is worth a thousand words" as an ancient Chinese saying, attaching the connotation of enhanced wisdom to the 1920's advertising principle [118]. Vision and comprehension of information are indelibly linked in our culture and language. To indicate understanding, we say "I see", getting at the details, we "bring it into focus", and removing ambiguities makes the data "clear" [25]. Information scientists are interested in more than just the intuition that visualizations add value and should endeavor to understand what makes them effective.

Visualization can harness the power of the human perceptual system to quickly identify structures and substructures within large amounts of data. Research in visualization seeks to provide design guidelines that help us to harness this technology for efficiently and effectively communicating and exploring our ideas. Until recently, the process of a graphic designer was laborious and the creation was static. The viewer was therefore limited to the role of a passive observer of a pre-formulated, explicit message. Advances in computer graphics mean that more complex displays can be produced and manipulated with minimal effort, opening the door to highly interactive representations of visual data. This newfound interaction allows the viewer to formulate, explore and validate hypotheses about the data, introducing a whole new dimension to the power of this tool: "It gives the user not only an eye for the data but also an opposable thumb" [29]. However, many questions remain about designing displays to support the information needs of the viewer. An important distinction lies between the ability to manipulate the pixels of a CRT and being able to capitalize on the strengths of human perception to convey a message. *Both* are essential for effective visualization.

## 1.1. The Challenge of Viewpoint Control

Many three-dimensional visualizations are characterized by the use of a mobile viewpoint that offers multiple perspectives on a set of visual information. Searching for and learning from objects in these displays requires both movement and orientation of gaze. While we routinely scan our surroundings using eye and head movements as we walk or drive through an environment, there are no corresponding *natural* movements of the viewpoint in desktop visualization systems. As a consequence it is more difficult to search for landmarks or generally understand our surroundings.

Viewpoint manipulation is a complex, yet critical task, often studied in the context of Virtual Environments (VEs). Hix concludes that the ability to manipulate the viewpoint to the appropriate positions is central to all tasks involving VEs [65]. Stanney concurs: "Without adequate means of moving about VEs, developers cannot expect to maximize human performance." [117]. To effectively control the viewpoint, the viewer must simultaneously manage the cognitive tasks of understanding the layout of the environment, and knowing where to look to find relevant information, along with mastering the physical interaction required to position the viewpoint in meaningful locations. Understanding these constraints and the ramifications of human capabilities give rise to several unresolved issues in the development of effective viewpoint control interfaces.

A major problem stems from the isolation of the visual link to virtual environments. When interaction with the target environment is limited to the visual channel, there is a breakdown of perceptual modalities, as well as a lack of important vestibular and proprioceptive cues. This impairment at the perceptual level leaves the viewer prone to numerous well-known, operational errors, including disorientation, degradation of situation awareness, failure to recognize hazards and simply overlooking relevant information [37, 86]. Accepting that remote perception will never match direct perception, the objective of VE systems should be to achieve *functional presence*. This occurs when the viewer receives enough cues to maintain situation awareness and successfully conduct operations in the remote environment [140]. Unlike the conventional understanding of “presence” [139], functional presence does not require operators to have the sense that they actually are situated at the remote location, only that they can accurately process the data that they are afforded. The ability to achieve functional presence may be promoted or hindered, depending on the control opportunities that are at the disposal of the operator.

Self-determination has been shown to be a major factor in learning about virtual environments. Peruch et al report that viewers who actively control the viewpoint are able to quickly gain a more accurate understanding of the environment [100]. In many cases however, a viewer is faced with the demanding task of simultaneously manipulating six viewing parameters: position (X,Y,Z) and orientation (Yaw, Pitch and Roll). This is not a trivial assignment and the effort applied to manipulating the viewpoint may compete with the task of extracting relevant information from the environment. Arthur points out that “When interaction becomes highly attention demanding, memory for the present location frequently decays, with the result that the individual becomes lost in space.” [2]. One way to facilitate viewpoint control is to develop and refine conceptual models that map controls to operational tasks. Several dominant metaphors for viewpoint interaction have emerged, including “world-in-hand” “eyeball-in-hand” “walking” “flying vehicle”, and “gaze-directed steering” [17, 132]. Despite the popularity of these techniques, successful navigation is also often dependent on adopting sophisticated strategies such as acquiring survey views, or moving is structured patterns [14]. Even with these strategies, there is still plenty of room for error; Goerger et al find that viewers in information-rich VEs are particularly susceptible to superfluous data and are easily distracted [54].

In fact, facilitating viewers' ability to extract relevant information from the surrounding context represents a challenge that is common to the design of all visual displays. This is commonly known among visualization researchers as the "focus+context" problem [25], and has been well researched in the domain of 2D displays. Many approaches to this problem allow the viewer to distort or magnify a portion of the display to allow for closer inspection [73]. Other methods attempt to provide visual cues to highlight or augment specific areas of interest [144]. While these approaches work well if the important features are always within the field-of-view, they may not be suited to large VEs where parts of the model can be occluded or otherwise out of sight. Furthermore, deliberately introducing distortions invalidates the spatial context and may impair the viewer's ability to gain accurate understanding of the overall configuration of the environment. This suggests the need for new tools or methods of interaction to actively help focus viewers' attention on relevant landmarks, while allowing extraneous features to be dismissed [54].

To ensure that relevant imagery is presented, researchers have attempted to develop mechanisms that provide guided tours through the environments. Several approaches have turned to the art of cinematography to come up with methods for dynamically creating animations that automatically steer the viewpoint through intricate three-dimensional displays [45, 61]. While this approach can be useful to highlight important features to naïve viewers, participation in the simulation becomes completely passive. Without direct user engagement, learning from the simulation can be significantly impaired [100]. Moreover, many of these techniques require a priori knowledge of the environment, a condition that cannot always be satisfied, especially in real-time visualization systems.

The numerous methods that have been developed for viewpoint interaction tend to cater to one of these two extremes: simplified controls or directed presentation. However, control of the viewpoint in 3D visualizations does not need to be an all-or-nothing scenario. A new objective should be to develop hybrid systems that share the burden of viewpoint control or allow viewers to toggle between user-controlled and guided modes. Very little attention has been paid to the idea of partially automating viewpoint control to promote a supportive, yet unscripted exploration of a 3D visual display.

A study of assisted viewpoint interaction can be subsumed by the broader field of recommender systems whose objective can be broadly defined as: "providing suggestions that effectively prune large information spaces so that users are directed toward those items that best meet their needs and preferences." [23]. When applying this definition to visual information systems, it is important to note that the "large information space" refers to the number of viewing options, not necessarily the size of the virtual environment. Recommender systems are often studied in the context of hypertext documents comprised of a set of discrete nodes. Travel to the different nodes is accomplished by simply selecting the desired node from a finite collection. Acquiring information from VEs is typically not that simple; there are often numerous viewing locations compounded by multiple orientations that can lead to a practically infinite array of viewing options, even for the smallest representation. This exacerbates the problem of recommending specific views – which are ideal? Moreover, it may be easy to assess a recommendation on a practical basis – "is the desired information in view?" – but aesthetic preferences may be harder to quantify. If the system is unable to directly present an absolute "best" view to the information, viewers might benefit from a system that allows them to cooperatively arrive at a view that meets their current needs.

## 1.2. Research Statement

The goal of directing the flow of relevant data to an individual is a central problem to the field of Information Science. In fact, this is the very definition of what it means to “inform”. **The primary objective of this research is to assess the effectiveness of interaction methods that *actively* support the information process through the visual information systems**. A general model of “Assisted Viewpoint Interaction” might dictate that the viewer is only directly responsible for a subset of viewing options, while secondary viewing parameters are restricted or possibly automated. This means that in addition to reducing the control space that the viewer must manage, recommendations can also take an active role. By automatically adjusting the viewpoint, the system can draw attention to features that the viewer might overlook if left to act alone.

At the same time, the bane of any recommendation system is the possibility of inappropriate or unwanted suggestions. This problem may be magnified when applied to viewing parameters; if viewers are out of synch with viewing recommendations, they may be forced to look at irrelevant areas of a display. Considerations must be taken to minimize these intrusions, lest they interfere with the user experience and negate the potential benefits. **Therefore, the second objective of this research is to understand the balance between user autonomy and directed presentation of information with regard to viewpoint assistance.**

## 1.3. Research Components

**Design Considerations for Assisted Viewpoint Interaction**- The first segment of this dissertation will outline the major issues that need to be addressed in the design of assisted viewpoint interaction. This framework serves to partition the design space according to various perspectives. From a user-centric approach, the designer must consider viewer interests and objectives for using visual information systems. From a data-centric perspective, there are certain constraints on the types of recommendations that can be made and how they are generated. There are also numerous interface options that ultimately dictate how the viewing experience can be better informed.

**Assisted Viewpoint Control System Development–** While each of the interface options described in the design framework has its appropriate role, this research is particularly interested in further exploration of interfaces that actively redirect viewers' attention while accommodating their independence. Brooks asserts that users have an intimate knowledge of their needs and objectives that can never be completely modeled or fully passed to the recommendation system [20]. The same holds true for viewers interacting with visualizations and thus they should remain dominant with respect to viewpoint control. Therefore, this module will focus on the iterative design and evaluation of attentive navigation, a specific class of assisted viewpoint interaction aimed at those criteria. This work will be conducted in two phases. First, a set of baseline studies looks for the benefits and drawbacks of an initial design of attentive navigation that does not factor in coordination between the human operator and the automation. This will allow an assessment of the impact of active viewing recommendations on the user experience. The second phase attempts to improve the design by suggesting formalized coordination protocols that will enhance the benefits or mute the shortcomings of the assisted viewing system. These iterative proposals are implemented and empirically evaluated to assess their usefulness.

### 1.4. Overview

This chapter has provided a high level motivation for this dissertation research and outlined the objectives sought. Chapter 2 presents some foundational literature that grounds this research in the Information Science discipline and further exposes the issues surrounding viewpoint manipulation. Chapter 3 addresses the first research component synthesizing a framework around the types of assistance, generation of recommendations and methods for informing the viewing experience. Chapter 4 introduces the details of attentive navigation and presents the results of three experimental evaluations, an iterative redesign and an additional evaluation.  Chapter 5 summarizes the findings of this research, assesses the potential impact and speculates on additional research that would support this topic.

# 2. Foundations

## 2.1. 3D Visualizations as Information Systems

The term “visualization” has traditionally referred to an activity that occurs within one’s head: “The formation of a mental image” [86]. However, Ware points out that the term is evolving to accommodate the “tangible, graphical representation of data” [129]. This paper adopts the notion of a visualization as a tool, similar to McCormick’s view: “Mechanisms by which humans perceive, interpret, use and communicate visual information” [84].

### 2.1.1. Visualization Mechanics

Several researchers have proposed models that attempt to classify how visualizations empower the viewer from an operational perspective [25, 78, 114, 129, 143]. Some dissect visualization into components of a communication process; others explore psychological models of human perception and cognition to understand how graphics are used to transfer information. Each of these frameworks sheds some light on the intents, effects and mechanics of visualizations contributing to a broad understanding of the topic. A review of these frameworks reveals that despite differences in scope, terminology and approach, there is considerable overlap in the conceptual coverage.

Spence [114] explores visualization as a process – a pipeline of operations that carry raw data to an analyst for the purpose of creating or influencing a mental model. The first step is selection and graphical encoding of the information that needs to be conveyed to the viewer. Although numerous encoding formats are discussed, Spence does not appraise them because “the value of a particular technique depends very much upon the application domain, the user, and the task being performed”. The next stage in Spence’s model is the presentation of the data. This has to do with the organization or layout of the graphical elements – when arranged properly, he argues, trends and relationships may be revealed that were otherwise obscured. The final stage of his model allows for the user to re-arrange, interact with and explore the display. While Spence is initially loath to admit the role of computer technology in the visualization process, he does acknowledge its importance to the final manipulation stage. He regards this step as a critical part of the model, worthy of development: "a great deal of effort has been invested in the invention and implementation of interactive visualization tools that harness this potential". This process, and the emphasis on an interaction loop are echoed by Cheng et. al. "All but the most trivial [visualization] problems require iterative cycles of (i) visual interpretation of the external diagrams, (ii) internal recognition of the applicable operators, (iii) modification of the drawing and (iv) further interpretation of the new diagrams."[28].

Another approach to understanding how visualizations function is to examine the information tasks that they are capable of supporting. This approach has been adopted several times, most notably by Roth and Mattis [103], Wehrend and Lewis [133], and Zhou and Feiner [143], producing the evolving taxonomy of tasks documented in Table 1.

**Table 1: Task Models of Visualization**

| Roth & Mattis[103] | Wehrend & Lewis [133] | Zhou & Feiner [143] |
|---|---|---|
| | Associate | Associate |
| | Background | Background |
| | Categorize | Categorize |
| | Cluster | Cluster |
| Compare | Compare | Compare |
| Correlate | Correlate | Correlate |
| Distribution of Values | Distinguish | Distinguish |
| | | Encode |
| | | Emphasize |
| | | Generalize |
| Look-Up Value | Identify | Identify |
| | Locate | Locate |
| Index | Rank | Rank |
| | | Reveal |
| | | Switch |

Zhou and Feiner further categorize the tasks into three underlying principles that explain how graphics can be used to communicate. *Visual organization* describes how the perceptual system influences the ways that people organize the world and perceive it as a whole. *Visual signaling* refers to the details of how the data is decoded to infer the meaning from a display. *Visual transformation* addresses methods for transitioning attention of the viewer and adapting to changes in the display.

Larkin and Simon address the benefits of visualization from a cognitive science perspective [78]. By comparing how physics problems could be solved using diagrams versus textual approaches, they suggest that there are two major benefits to visualization: computation and searching. They propose that analysts can substitute perceptual inferences for more complex logical deduction or mental arithmetic, especially with respect to distance, size, spatial coincidence and color comparisons. Likewise search time can be reduced by spatial arrangement – grouping like objects together can allow the analyst to quickly identify outliers. This locational indexing often makes diagrams “more effective than informationally equivalent sentential representations” [28].

Ware is a strong advocate of the perceptual benefits of visualization, describing visualization as a “science of sensory representations” [129]. His model describes the advantages of visualization in terms of how the human perceptual system interacts with characteristic features shared by many graphics: 1) Compactness – the ability to comprehend huge volumes of data from a relatively small space. 2) Emergent properties are visible – ranging from detecting patterns to identifying outliers. 3) Large-scale and small-scale features of the data are accessible, allowing specific data points to be put into context. Finally, he adds that visualization is a powerful tool for hypothesis formation, allowing the viewer to quickly see the impact of changes made to the data.

Card, Mackinlay and Shneiderman [25] synthesize the literature to provide a useful taxonomy for the effects of visualization. According to their work, visualizations amplify cognition in six major ways: 1) Increasing the memory and processing resources, 2) Easing the search for information, 3) Facilitating the ability to recognize patterns, 4) Enabling perceptual inference, 5) Harnessing perceptual monitoring abilities, 6) Encoding information into a manipulable medium.

These frameworks can be collapsed into three major mechanisms for visually augmenting cognition: Perceptual Enhancements, External Cognition and Interactive Manipulation. Table 2 maps the frameworks discussed above onto these principles.

**Table 2: Mechanisms for Visually Augmenting Cognition**

| | **Perceptual Enhancements** | **External Cognition** | **Interactive Manipulation** |
|---|---|---|---|
| Spence | Presentation | | Manipulation |
| Cheng | Visual Interpretation | Recognition | Modification |
| Zhou & Feiner | Organization | Signaling | Transformation |
| Larkin and Simon | Search; Computation | Search; Computation | |
| Ware | Emergent Properties | Compactness; Levels of detail | Hypothesis Formation |
| Card, Mackinlay & Shneiderman | Enhanced recognition of patterns; Perceptual inferencing; Perceptual monitoring | Increased resources; reduced Search | Manipulable medium |

There are numerous psychological theories of perception that can be leveraged in the construction of visualizations to facilitate decoding and enhance the transfer of information. Central to many of these theories is the existence of a short-lived 'Iconic Buffer' or 'Sensory Store' proposed by Sperling [cited by 1] that serves as the first stage of this decoding process. By understanding what this buffer contains, and how it is accessed and organized the design of the display can be influenced to allow "quick perceptual judgments"[77] and tap into the power of the "highly parallel human visual system to replace more cumbersome serial logical inferences" [28].

External cognition is the second major way that visualizations can augment cognition. Instead of relying on knowledge that is completely internal, people are able to quickly integrate information from the environment to enhance their overall comprehension. By actively creating external representations, working memory loads and other cognitive demands may be reduced. This can be easily demonstrated with an informal study comparing performance time for completing multiplication problems mentally versus with pencil and paper [25]. Moreover, external representations are able to "constrain the kinds of inferences that can made about the underlying represented world" [105]. This allows some relationships to be encoded visually, eliminating the need for extra cognitive processing.

To a certain degree, the first two mechanisms, perceptual enhancement and external cognition can be subsumed by a study of interaction; they act as the words in an interactive discourse. For instance, as a result of manipulating the display, perceptual cues can be used to respond to the viewer's request to make a subset of the data distinct [109], or the computer may assist with completing a challenging mental tasks, such as rotating 3D objects [107].

Card emphasized the importance of interactive manipulation in the forward to a recent monograph: "Thinking of visualizations as insightful pictures leaves out the important role of manipulation, of operations done to the data in order to tease out yet other revealing images"[29]. Even Bertin, whose seminal works predate highly interactive graphics systems recognized the need for interaction: "A graphic is not 'drawn' once and for all; it is 'constructed' and reconstructed until it reveals all the relationships constituted by the interplay of the data… a graphic is never an end in itself; it is a moment in the process of decision-making" [13].

Haber and McNabb [55] describe the opportunities for interaction at various stages of the visualization process in terms of three transformations, as shown in Figure 1. The viewer can potentially interact with the visualization to influence each of these transformations, affording the opportunity to gain new insight in to the data that is to be displayed. The process begins as raw data is converted to derived data through *enrichment*. This process fits the data to a desired model and may include various data manipulations such as smoothing, filtering, interpolating, or even making hypothetical alterations. The second stage, *mapping*, converts the dataset to an abstract visual object (AVO). In this step, attributes of the data are converted to corresponding geometric features according to a set of classification rules. By adjusting these rules, the viewer can cause distinctions in the data to become perceptually obvious. This step establishes the representation of the data, but not necessarily the final visual appearance of the display. *Rendering* ultimately converts the AVO to a displayable image. This involves computer graphics techniques such as view transformations (rotation, translation, scaling), hidden surface removal, shading, shadowing, etc. These final transformations alter the perspective on the display, contribute to the overall appearance and more importantly may even dictate what is or is not visible.

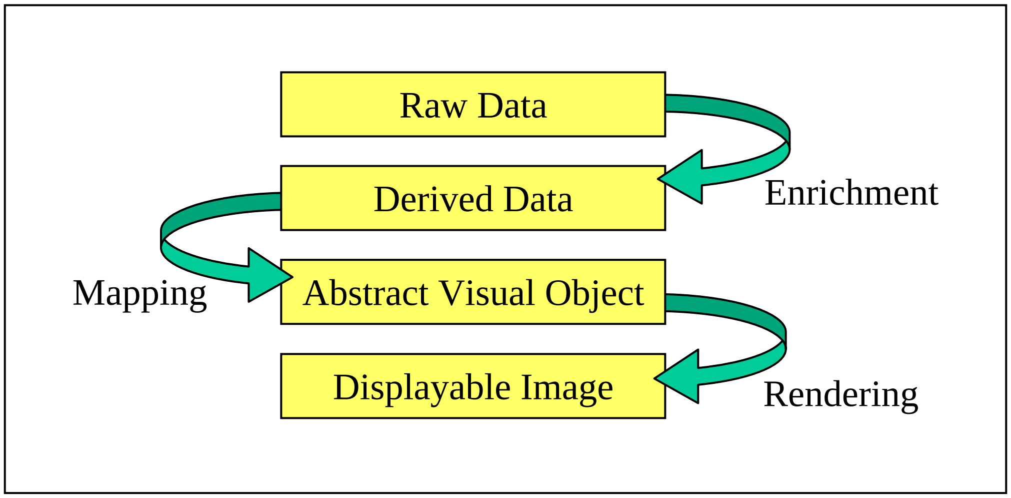



**Figure 1: Haber & McNabb's Visualization Process**

An alternative model of interaction characterizes interactivity in terms of interlocking feedback loops [129]. The *visual-manual control* loop addresses physical interaction with the system by which “objects are selected and moved using basic skills of hand-eye coordination”. The second level is the *view refinement and navigation* loop. In this stage, the viewer manipulates the display to explore the visual data space in order to better understand the underlying relationships. These two loops are bound together by a third loop, the *problem-solving* loop, which integrates the viewer’s internal representations with the external visualization to motivate actions to be taken.

It is possible to combine these two schemes by examining how these interactions change the representation. Working within the visual-manual control loop, one can enrich the data or adjust mapping rules, to make *structural* alterations. Likewise view refinement and rendering determine the viewer’s perspective on the display; changes at this level result in *viewpoint* alterations. This dichotomy is also reflected in Shneiderman’s work on interaction. Shneiderman is well known for his work on Direct Manipulation interfaces, which allows elements of the display to be manipulated rapidly, incrementally and reversibly, with immediate feedback [108]. Clearly, this kind of interaction is useful for structural alterations. With regard to visualization interfaces, Shneiderman also impresses the importance of the viewpoint manipulations through his visual information seeking mantra “Overview first, Zoom and Filter, then Details on Demand”[110]. Table 3 summarizes the interaction methods for visualization.

**Table 3: Visualization Interaction Methods**

| | **Structural Alterations** | **Viewpoint Alterations** |
|---|---|---|
| Haber and McNabb | Enrichment, Mapping | Rendering |
| Ware | Visual-Manual Loop | View-Refinement and Navigation |
| Shneiderman | Direct Manipulation | Visual Information Seeking “Mantra” |

Navigation and adjusting the viewpoint in virtual environments have been identified as having a profound impact on all other user tasks [64]. Viewpoint interaction is critical in the sense that it allows the viewer to determine what portion of the data is actually displayed; adjustments can be made to find and focus on particular regions of interest. However, the benefits of being able to actively manipulate the viewpoint extend beyond direct information foraging. Gabbard and Hix suggest that the ability to view scenes and objects from many different angles fosters an overall understanding of the environment [51]. Moreover, viewpoint motion assists with depth perception and may also contribute to an understanding of the environment's global configuration.

### 2.1.2. 3D Representations

The previous section provided a broad explanation of how visualization can be an effective tool to augment cognition. Shneiderman argues that the type of data that is being represented heavily shapes visualization benefits[110]. Specifically, he characterizes seven types: 1-, 2-, and 3-dimensional, temporal, tree, network and multi-dimensional (n>3). Each of these representations has it's own benefits and drawbacks, and is the focus of separate sub-fields of study. For example, presenting 3D information taps into the familiar arrangement of space, but its representation is often a projection onto a 2D medium (paper, computer monitors, etc), potentially introducing perceptual difficulties. The ongoing explosion in computer graphics technology is making it easier to manage the numerous parameters that contribute to the illusion of 3D.

Using 3D displays is not without its detractors. Echoing the thoughts of Ivan Sutherland and Fred Brooks, Hinckley counters that "People may experience 3D, but they do not innately understand it" [62]. He considers the difficulties of building a stone wall to convey problems that the general public might have with 3D interfaces. Spring observes, "If virtual reality represents the future of computing, that means that we will need to revert to counting on our virtual fingers" [115]. We recognize this quip as a hyperbole, but also note its nugget of truth: blindly defaulting to 3D representations is not a good idea. It may seem obvious that data that is inherently 2-dimensional should not be represented with a 3D display, however this occurs with an alarming frequency. Tufte points out numerous examples of how this kind of display violates his design principles by increasing the data-to-ink ratio and creating "Chart-junk" [122]. Ware blames the temptation to use unnecessary 3D graphics on its novelty and growing simplicity and observes similar trends with inappropriate font changes and gratuitous color as those technologies matured [129].

Heeding these admonitions, there are numerous practical applications of 3D technology. Advocates claim that natural experiences with the real (3D) world will transfer into synthetic (3D) visualization displays [25]. This is the basis for simulation-based training and virtual reality. The 3D visual component of the simulation allows the experimenter to quickly assess the progress or state of the simulation. It may also simply be the case that the data represents an inherently 3D object that is not directly viewable. When surveying the literature, three data attributes quickly emerge that characterize this rationale:

- **Disparate Scale**: Objects can either be too small or too large to be viewed naturally. Examples include: (micro) molecular inspections, and (macro) geographic phenomena
- **Prohibitive Region**: Data represents an environment that is inaccessible, hazardous or costly. Examples include: medical imaging, undersea or extra-planetary exploration, or urban search and rescue activities.
- **Speculative Region**: The object, or parts of the object in question do not exist in the physical world. Alternatively, this kind of display can also be subtractive, contemplating the absence of an exiting object. Examples include: computer-aided design (CAD), archeological reconstruction, and urban planning.

### 2.1.3. Levels of Abstraction

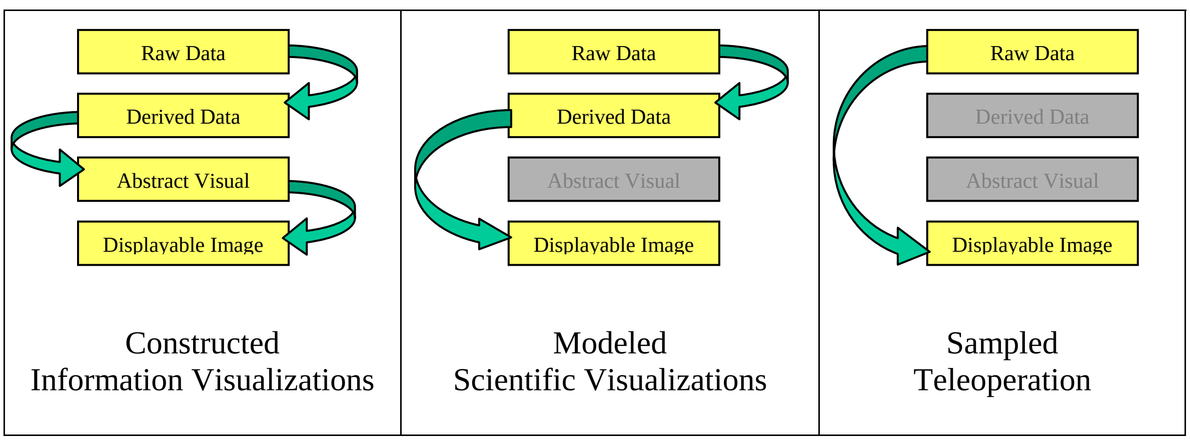
Another important aspect of visualization is the level of abstraction between the raw data and the visual representation. Kosslyn categorizes symbolic displays into four types, based on the relationships among the constituent data [75]. *Graphs* represent the relationship between two or more scalar variables, where position is determined by a pairing operation. *Charts* identify discrete relationships among entities using lines, enclosure and relative position. *Maps* convey inherently spatial relationships by pairing labels with semantically meaningful locations. Finally, *Diagrams* provide schematic pictures of existing objects.

Inherent to Kosslyn's taxonomy is the notion that data represented in visualizations can be abstract or, in the case of many maps and diagrams, constrained by real-world constructs. This is an important distinction that forms a primary axis in the field of Visualization. Data that are inherently spatial fall into the category of *Scientific Visualization*, while data that are more abstract are represented by *Information Visualization*. Munzner comments on this unfortunate naming convention, observing that there are abstract representations of scientific data and information is certainly not absent from inherently spatial displays [91]. Despite the naming issues, the distinction between the two categories is very important. Given their abstract nature, information visualizations are fully *constructed*, i.e. the designer of the visualization decides how to represent elements of the data (i.e. with a red square or a blue triangle). In contrast, traditional scientific visualizations are produced by mapping data onto an existing spatial substrate that must be *modeled.* Under these conditions, the designer is significantly limited in selecting a representation to convey information.

Casting a broad net, a third type of visualization might be considered given the novel facility of remote computer imaging. Instead of a computer-generated model, the spatial substrate of the visualization may be *sampled* from the environment and directly presented to the viewer. While this new category might be considered a part of scientific visualization, it carries the important distinction that the designer cannot modify the fundamental structure, and thus merits separate discussion. Alternatively, this kind of display may be thought of as separate from the field of visualization entirely, and is often studied under the label of Teleoperation. However, Milgram observes compelling parallels in terms of cognitive demands, operation, and overall human experience, suggesting substantial overlap between the two fields and warranting its inclusion in this discussion [88].

The level of abstraction influences the opportunities for interaction. Table 4 offers a revised view of Haber and McNabb's interaction model in light of the different levels of abstraction. Since modeled visualizations rely on inherently spatial data, the need for abstract visual objects is diminished. Likewise with sampled visualizations, we see that raw data is directly presented as a displayable image. This leaves the rendering stage – responsible for viewing transformations – as the primary interaction common to the three types of visualization[1].

**Table 4: The Visualization Process by Levels of Abstraction**




In terms of 3D viewpoint interaction, research has been conducted in all three of these sub-fields. However, in an attempt to synthesize the breadth of all three classes into one discussion, some

[1] It should also be noted that Telepresence also frequently requires a form of structural alterations – although this interaction is fundamentally different because it physically alters the real environment which results in a change in the raw data.

caveats must be understood. First, of the three types of visualization reviewed, the use of 3D representations for information visualization is the most contentious. Since the entire display must be constructed from abstract data, the author has extreme flexibility in terms of its representation. Alternative representations such as multiple 2D displays or simply encoding the data using retinal variables other than position may be able to convey the desired message just as well. While the techniques for 3D viewpoint interaction discussed in this paper can apply to information visualizations, adopting a 3D representation is largely optional, and ramifications for perceptual decoding and interaction must weigh in. Despite this concern, several 3D information visualizations have already proven successful, especially for managing large network [98] and hierarchical [102] information.

From the other side of the spectrum, interaction with sampled environments is highly constrained. Not only are the contents bound by reality, but the movement of the viewpoint must also correspond to the physical movements of a real camera. One of the purported benefits to exploring an artificial environment is that constraints of the physical world can be abandoned. For example, viewers can instantaneously teleport from one spot to another, or fly above the scene to gain a global perspective. Such actions may not be available in traditional teleoperation environments. However, effective techniques may inspire new designs and novel approaches. For example, Thomas suggests deploying a camera on a tethered balloon to acquire a birds-eye view of the environment [120].

Finally, it should be noted that these levels of abstraction do not denote discrete containers that neatly classify visualizations. Rather, they should be viewed as cluster points on a continuum from highly abstracted information visualizations to direct presentation of sampled environments. Some visualization techniques, such as Augmented Reality [4] blur the boundaries by registering auxiliary images over a sampled environment.

## 2.2. Viewpoint Manipulation

In many computer graphics packages, the viewpoint is described using a virtual camera analogy. The viewing region and visibility of the data is defined by settings on the camera. Consistent with the metaphor, the virtual camera can be manipulated just like a real camera – it can be located at different positions, it can be pointed in the direction of specific objects, zoomed, focused and so on.

While numerous means of viewpoint manipulation exist, including magic lenses [126], distortion techniques [73], and zooming [11], camera movement is the most common. There are six parameters that express motion of the camera or viewpoint with respect to a 3D environment. The position of the camera is described by three degrees of freedom – namely, the displacement of the camera along the three axes – X, Y and Z. The orientation of the camera accounts for the remaining three degrees of freedom. The Yaw, Pitch and Roll of the camera describe the magnitude of rotation around the three axes.

Moving the camera is usually expressed in terms of three objectives: Exploration, Search and Maneuvering [14]. The goal of exploration is to advance viewers' spatial knowledge of the environment independent of any specific targets. Searches seek to determine and track to the location of a certain object or class of objects. A primed search occurs when the location is known, in contrast to naïve searches where the operator has no a priori knowledge of the target's location [39]. Finally, maneuvering occurs when a specific viewpoint or series of viewpoints is required for a target object. This activity, also known as inspection, usually involves short, precise adjustments within a localized boundary. While these tasks are considered mutually exclusive, they are frequently compounded into sequences [40]. For example, operators engaged in an exploration may need to maneuver to identify a discriminating feature of an object that has caught their attention.

To explain how the viewer manipulates these camera parameters, the discussion focuses on three factors. The *Frame of Reference* dictates the camera's position and movement relative to the objects in the environment. A study of *Input Devices* provides insight to the viewer's physical interaction with the system. Finally, *Motion Control Metaphors* explains how the viewer's actions are interpreted by the system to affect the motion of the camera.

### 2.2.1. Frames of Reference

One of the earliest classifications of viewpoint manipulation for virtual environments divides techniques between "Eyeball-in-hand" systems and "World-in-hand" systems [130]. This classification highlights the distinction caused by applying transformations to different coordinate systems.

Applying transformations to a coordinate system that is centered on the camera, as shown in Figure 2, creates the sense that the viewpoint is moving through a scene. This egocentric motion has been shown to provide a sense of immersion, leading to a better understanding of the configuration of large environments [112].

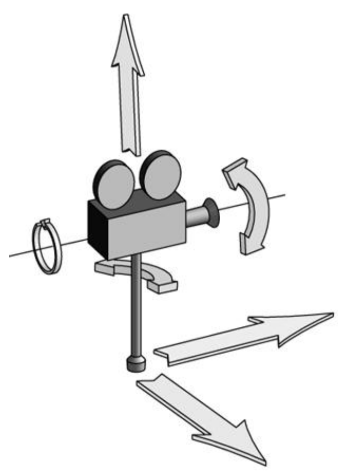


**Figure 2: Egocentric viewpoint adjustments**

Alternatively, "World-in-hand" techniques leave the camera fixed while applying the transformations to an arbitrary point in the model, as shown in Figure 3 [130]. This type of manipulation gives the viewer the sense that they have grabbed a portion of the display and can drag or spin it according to their wishes. This frame of reference has been shown to be effective for manipulating closed objects, but proves challenging if the viewpoint is to move among objects [130]. For this reason, this frame of reference is also sometimes referred to as an "object-view"

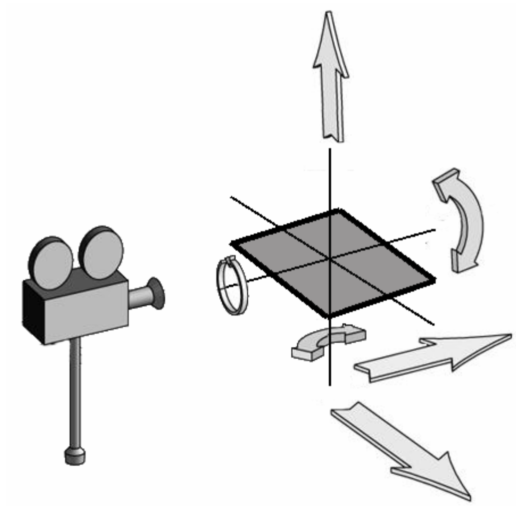


**Figure 3: Exocentric viewpoint adjustments**

These two frames of reference offer a tradeoff between local control and global awareness. A hybrid approach can be achieved using the metaphor of a tethered camera [88]. In this condition, the viewer controls the actions of an avatar in the environment, and the virtual camera trails behind, constrained by the motions of the avatar. This approach allows the viewer to better understand the motions through the environment by providing more contextual information. These additional cues come at the expense of a sense of presence – since the viewer is necessarily detached from the object that is exploring the environment. The relationship between these frames of reference are highlighted in Figure 4.

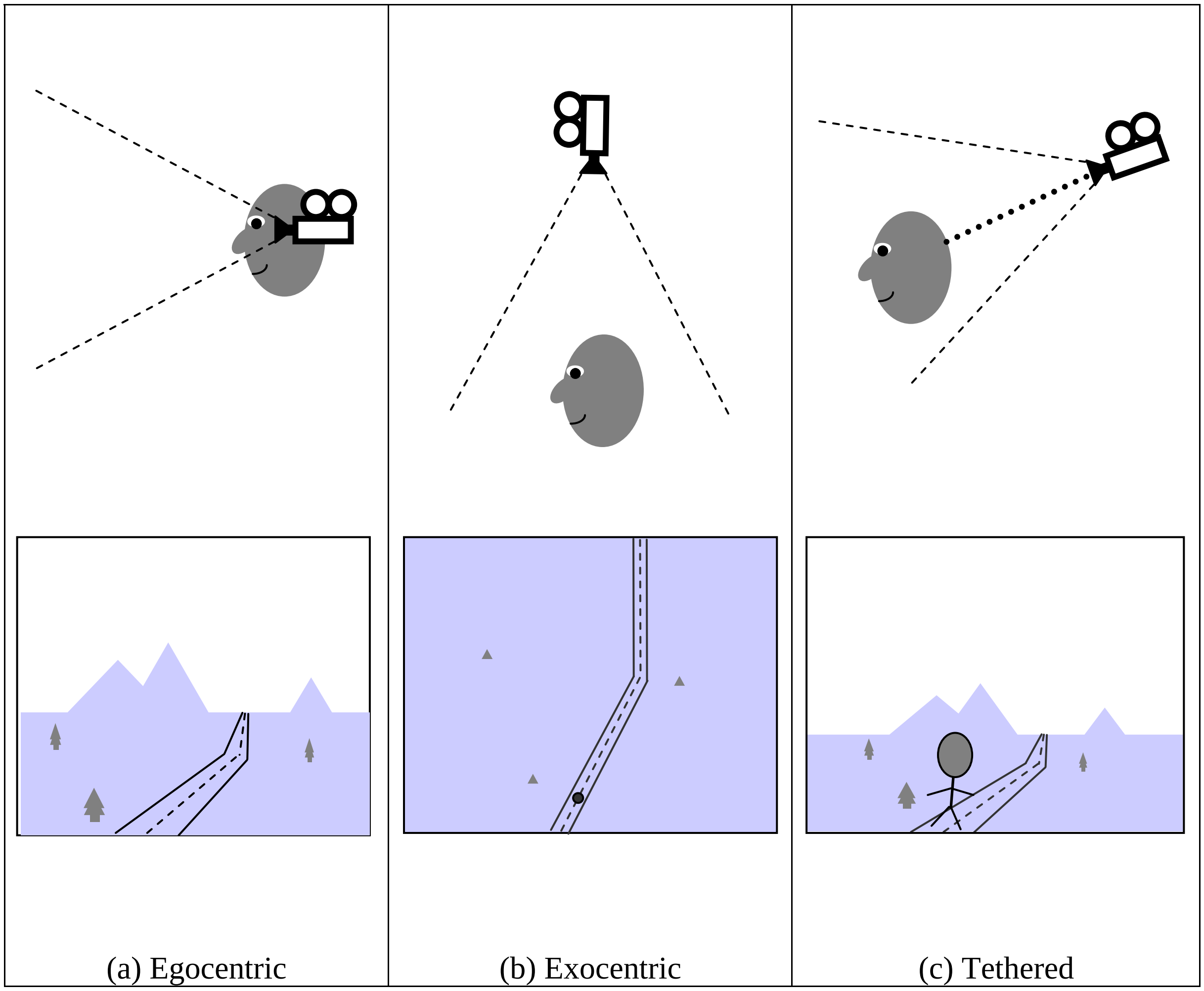


**Figure 4: Viewpoint Frames of Reference**

Another hybrid approach is to simultaneously provide multiple viewpoints. This is the basis for the "Worlds-in-Miniature" concept [118]. Embedded in the egocentric view is a smaller exocentric view that allows the viewer to obtain a bird's-eye view of the entire environment. Using this macro-view, the user can understand the global configuration of the environment, while making minor adjustments using the egocentric view. Plumlee refined this technique to ensure that the relationship between the multiple viewpoints could be easily understood. [100]. As shown in Figure 5, this design consists of binding an embedded egocentric viewpoint to an exocentric viewpoint via connecting lines[2]. Moreover the viewing frustum that defines the first person perspective is dynamically overlaid on the map view through a proxy, allowing the viewer to understand the extent of the world being viewed.

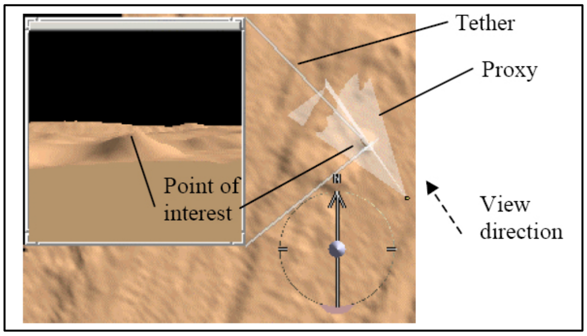



**Figure 5: Connecting Multiple Viewpoints [100]**

### 2.2.2. Input Devices

Input devices provide the mechanisms by which the viewer communicates intentions to the display. Mine provides a useful taxonomy that divides techniques according to the type of input device used, Direct, Physical and Abstract [89].

Direct controls register the viewer's natural gestures and body movements as input to the visualization system. For example, the viewer may point, look around or even walk on a treadmill or ride a stationary bicycle to indicate their intentions to the system [14]. Direct controls are often considered very naturalistic because the viewer interacts with the visualization in the same way that they would with the real world and require little or no additional training in order to master artificial controls. However, with the benefits of natural interaction comes the expectation of real-time responsiveness. Any latency in the reaction of the system can have a severe negative impact on the user experience [14]. Unfortunately, the cost and availability of sensors that afford this kind of performance often make this kind of interaction prohibitive to casual users at this time.

---

[2] Plumlee calls these connecting lines a "tether". This should not be confused with the tethered camera described previously, as the second viewpoint is completely egocentric and does not include an avatar.

Physical controls introduce a layer of abstraction into travel operations. Instead of directly adjusting the viewpoint through repositioning our body and reorienting our gaze, these actions are accomplished by mapping these movements to a physical device such as a joystick, mouse, or game pad. While these controls do not represent "natural" interactions, they may nonetheless be familiar. For instance, a steering wheel is a very common example of a physical control.

In determining the effectiveness of physical controls, one thing that must be considered is the degrees of freedom that it offers. Six degree-of-freedom (6DOF) devices offer a one-to-one relationship between control options and viewing parameters. While numerous 6DOF controllers exist, they can be effectively classified into two groups with respect to viewpoint manipulation: Free-moving and Stationary [141].

Free-Moving devices, such as Ware's flying mouse, a.k.a. the "Bat" shown in Figure 6, rely on directly mapping positional changes in the device to positional changes in the display. This kind of mapping is often referred to as a zero-order control. While this kind of mapping is fast and intuitive, it does lead to several problems [141]. The most serious is that anatomical limitations can severely impede motion. For example, the range of motion in the human wrist makes it very difficult for the viewer to spin 360° in place. Likewise, a user may find it challenging to move forward if his arm is already fully extended. Unlike a regular mouse, which measures displacement on a planar surface, the device cannot simply be picked up and repositioned. A special "clutching" mechanism must be built into the device to allow it to be temporarily disabled while it is repositioned. By some accounts the clutch design is the most significant usability concern with this kind of device [62]. Another significant problem often reported with this kind of device is fatigue. The user often has to work with their arm extended and unsupported for lengthy periods of time without rest.

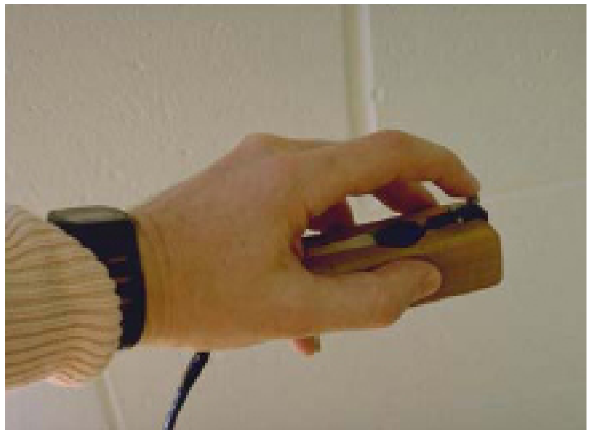


**Figure 6: Ware's Bat (Flying Mouse) [128]**

Stationary controllers, in contrast, register movement relative to a stationary base. Since there is a limited range of motion from the base, these devices tend to be rate-based or first-order controllers. This means that the displacement of device affects the velocity of the viewpoint changes in the environment. These stationary controllers overcome fatigue by allowing the operator to rest their arms on the desktop while operating. Moreover, the rate control mechanism obviates the need for a clutching component.

Despite some advantages over their free-moving counterparts, these devices have problems of their own. Many instances of these stationary controllers, such as the SpaceMaster and SpaceOrb shown in Figure 7 are isometric, meaning that they register force, but do not move by a significantly perceptible magnitude [141]. This causes the device to feel rigid and lacking in haptic feedback. This coupled with anatomical constraints can cause inadvertent collusion between translational and rotational operations [62].

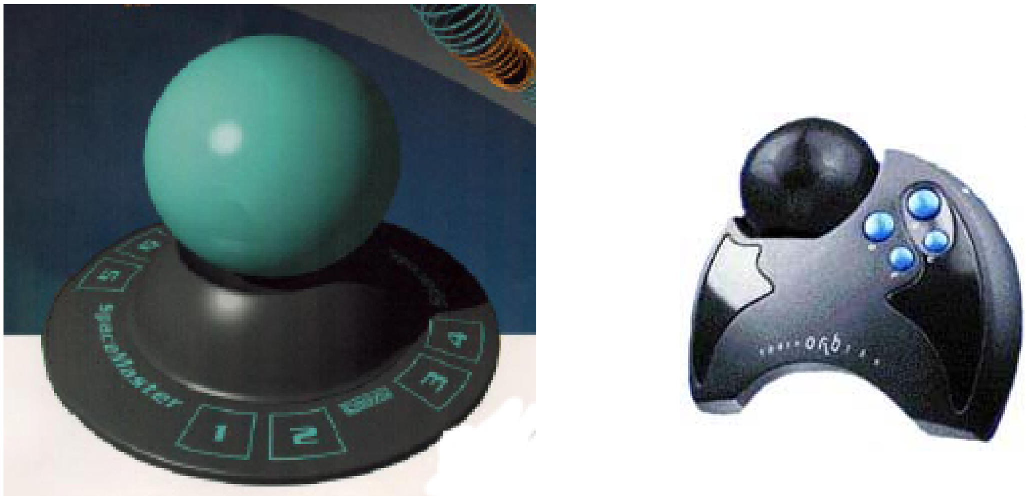



**Figure 7: SpaceMaster and SpaceOrb**

While six degrees-of-freedom are necessary to fully control motion in three dimensions, the cognitive overhead of operating these devices may be more than the average user can handle, leading to disorientation [2]. Furthermore, providing simultaneous access to all six degrees-of-freedom may not be consistent with the way that users operate in a 3D environment. Viewers tend to allocate attention separately to the translation and rotational degrees of freedom, alternatively switching between the two groups. [82]

While the cognitive and physical demands of using a 6DOF controller might be overcome with training, there is the practical consideration that these higher order input devices are not currently mainstream controllers. A highly practical application or radical new design may yet motivate a shift in standards, however in the interim, research has produced a rich literature of techniques that try to creatively map viewpoint controls to controllers that have lower dimensionality. These strategies can be classified into several categories based on how they reduce the needed degrees of freedom. Four major clusters emerge from this analysis: Overloading, Constraining, Coupling and Offloading.

*Overloading* – Extra degrees of freedom are achieved by modal operation of the device. Various combinations of control keys or button presses supplement the operation of the device to determine the mode of operation. While this technique is popular with CAD and modeling software, the increased cognitive burden of remembering the current state can negatively impact performance [8]. In addition to operational difficulties, Gabbard and Hix report qualitative demerits for this kind of modal interaction; users complained the constant need to toggle between modes was "frustrating and counter-productive." [51].

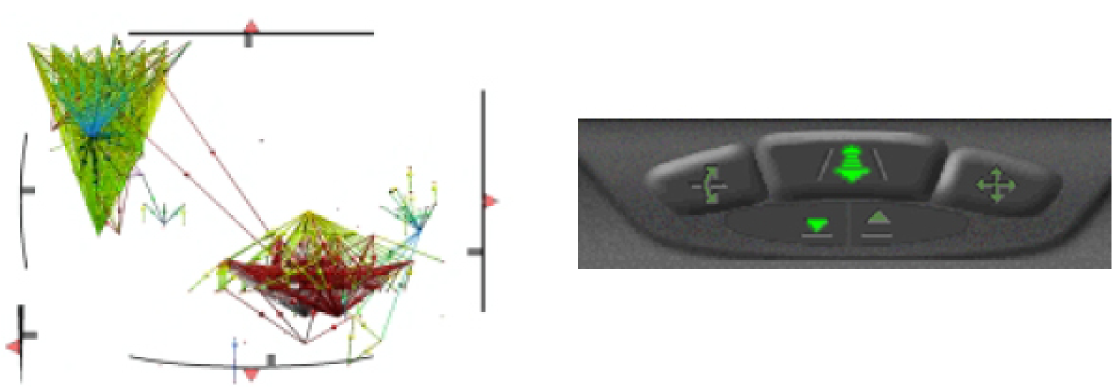
*Constraining* – Movement of the viewpoint is limited to certain operations; manipulations of other attributes are simply discarded. The most common example of constraining is to restrict motion to a ground plane, eliminating the need for vertical translation [130]. Roll is also frequently eliminated, especially in simulations of ground vehicles.

*Coupling* – This approach functionally binds one or more viewpoint attributes to the state of the others. The most common example is known as gaze-directed steering, in which the viewer's motion is determined by the direction they are looking [18]. Using this technique, the three positional variables can be condensed into a single control – velocity.

*Offloading* – This method cedes control of certain travel operations to an external source. These sources may include a pre-computed route or sequence, a collaborative operator or even a computer-driven agent.

The final layer in Mine's framework abstracts operations even further to Virtual Controls. This class of controls includes on-screen representations like those shown in Figure 8. This type of control offers a great deal of flexibility in that the interface can be completely customized to fit the need of the application. For example, Parker developed a collection of sliders that proved quite effective at providing a world-in-hand view for networked data (Figure 8a) [98]. However, these widgets would not have been appropriate for an application where the task was exploration of an architectural model where an egocentric frame of reference is needed. Instead one might consider virtual controls similar to the control panel provided by the Cosmo Player (Figure 8b) [106].

As with the other input mechanisms, several problems are associated with virtual controls. Since the control is completely synthetic, there is no haptic feedback that the control has been activated [89]. Although other senses could be used, a visual cue is often the only indication which virtual manipulator is in use. Obviously diverting visual attention to the control panel means that the viewer is not fully attending to the information display, potentially reducing the efficiency of the interaction. Another issue that is often raised with virtual controls is the problem of issuing simultaneous commands [141]. Coordinated movements can occur naturally with direct and physical controls, but it is more difficult to select more than one operation from a screen-based control panel. Finally, the use of virtual controls often takes up valuable screen real estate. Not only do these interfaces compete for the visual attention of the viewer, but they also limit the perceptual substrate that is available for presentation of information.



(a) (b)

**Figure 8: Virtual Controls: a) 3D Widgets (Sliders) [98] b) Cosmo Player control panel [106]**

### 2.2.3. Motion Control Metaphors

The use of metaphors in interface design allows the user to exploit prior knowledge and take familiar actions when dealing with a new technology [93]. Several travel metaphors have been identified as a means to reduce the cognitive load required to manipulate the viewpoint to meaningful perspectives [6, 19, 51]. These include steering, route drawing, target-based and discrete selection.

Capitalizing on the commonplace understanding of driving an automobile, steering is one of the most widespread metaphors for viewpoint interaction. The operator repositions the viewpoint through a combination of adjustments to the direction and speed of motion [89]. Typically this works in a sequential manner; the viewer selects a promising orientation, and then initiates movement, telling the virtual camera to “move forward”, until the desired viewpoint has been acquired.

Several alternatives exist for establishing the direction of travel, however, the most pervasive is known as gaze-directed steering [18]. Effectively, the direction of travel is determined by the orientation of the camera. Operationally, this approach is simple to manipulate; forward motion is always in the direction that viewer is facing. Its popularity is based not only on familiarity and simplicity, but also on its ease of implementation. In teleoperation activities, for example, this control technique is achieved by simply mounting a fixed camera on the front of the robot.

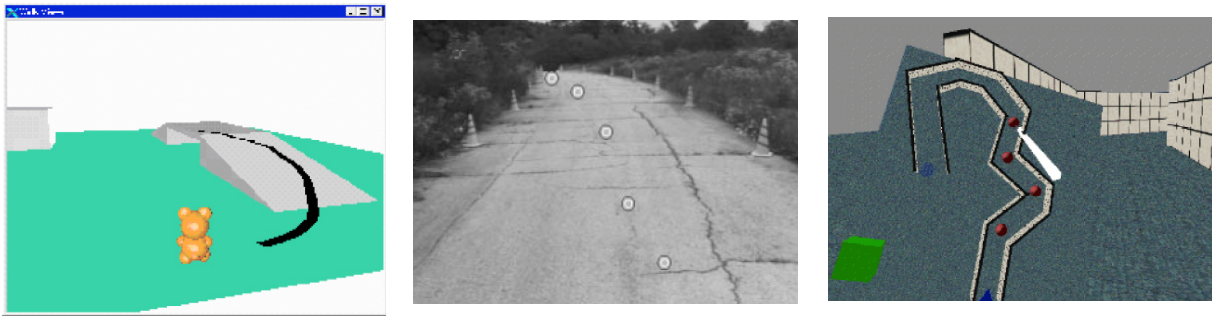
However, this highly intuitive technique may come at the expense of all but the most trivial inspections. Consider the task of looking at an object from all sides. Since the viewpoint always moves forward in the direction that the camera is oriented, there is no guarantee that the object of interest even remains in the field of view, increasing the chances that useful viewpoints may be overlooked or missed [41].

Alternative steering approaches that allow the direction of travel to be determined independent of the gaze direction are often called pointing techniques. These techniques allow the viewer to point in an arbitrary direction to establish the motion vector. While this affords the flexibility of subjective motion, it may be confusing for users to maintain the relationship between the gaze direction and motion direction [89]. Furthermore, without a very good understanding of the environment, it would be ill advised to spend much time navigating with this method, for the simple reason that the operator may not be able to see where they are going, risking collisions.

Another category of viewpoint manipulation techniques is known as route drawing. Instead of continuously adjusting the direction of travel, the operator specifies a route segment and then executes travel along that path. There are two major approaches to route drawing, defined by whether the path is drawn directly on the viewing scene or on an external map of the environment.

The path-drawing technique, described by Igarashi, allows the operator to draw a path directly on the viewing scene (Figure 9a) [70]. This technique factors in the geometry of the environment to map the route specified to a walking surface, and also uses this information to assist with obstacle avoidance and intelligent behavior with architectural features such as gates and tunnels. Kay's STRIPE system (Figure 9b), implemented a similar approach for a teleoperated vehicle [71]. Instead of the continuous line of the path drawing systems, however, Kay's approach used individual waypoints. The robotic vehicle attempted to reconcile the known ground-plane with the next waypoint to determine which direction to move the vehicle.

The other kind of route drawing involves the use of a map or alternative viewpoint. Bowman demonstrates how the path can be specified on a map of the environment (Figure 9c) [14]. Much like Kay's approach, a series of markers are placed on the map, indicating desired viewpoints. Upon execution, the system interpolates the gaps between the markers, moving the viewer to each of the specified viewpoints. The map techniques allow for larger navigation sequences to be programmed at a given time, however, they require knowledge of the entire environment that may not be available in real-time applications.



(a) (b) (c)

**Figure 9: (a) Igarashi's Path Drawing, (b) Kay's Stripe System, (c) Bowman's Map and Stylus**

The major benefit to this type of navigation is that it allows the viewer to maintain control over the position of the viewpoint, but other operations can be performed during motion [19]. Adopting this strategy allows the operator to focus on navigation issues at the beginning of a task, freeing up cognitive resources for other activities while the viewpoint actually moves. Another reason to consider these approaches is latency. Rendering delays or communication lags can cause serious problems if the operator were using a steering system that demands continuous adjustments [70, 71].

Target-based viewpoint manipulation metaphors allow the viewer to select features in the environment and adjust the viewpoint relative to their selection. This mode of navigation is particularly useful when the viewer needs to further inspect the details of an object. Two common scenarios where this kind of manipulation proves useful include assisting the viewer to get near target objects and allowing the viewer to manipulate the viewpoint to acquire multiple perspectives on the target.

Mackinlay notes that people often have difficulty successfully manipulating the viewpoint with respect to a desired target object [81]. He observed users either moved very slowly through the environment to avoid making mistake, or they move very quickly, often overshooting the target and then compensate for any mistakes. A third strategy has viewers moving in short quick pulses with pauses in between to assess the effectiveness of the last move. To address this problem, he proposed "Point-of-interest" (POI) navigation where the viewer could specify a target objects. When the viewer started moving toward the target, the distance that the viewer had yet to travel to the target logarithmically modulated the speed. This allowed the viewer to move quickly if they had a large expanse to travel, but afforded more control as they neared their destination. Thus POI navigation optimizes the travel time and minimizes the potential to overshoot the desired viewpoint location.

The second scenario allows the viewer to see a target object from multiple perspectives. This is often characterized by going into an "orbit" around the target object The specifications for VRML 97 [131] and its successor X3D [132], define a built in target-based mode for navigation: "`EXAMINE` shall provide the ability to orbit or spin the user's eyepoint about the center of rotation in response to user actions."

The final major category of locational viewpoint manipulation is Discrete Selection. This technique allows for the viewer to immediately jump to a new viewpoint. This class of techniques is often characterized as a form of target selection. However, there are three important distinctions. First, as defined above, target-based techniques allow the viewer to control the viewpoint relative to a particular target. In contrast, discrete selection only moves the viewer to a particular point of view; any subsequent motion is out of the scope of this metaphor. Second, the viewpoints available by discrete selection are not necessarily associated with a "tangible" object in the environment. For example, the viewer may select a viewpoint that “looks down the hallway”. Third, target-based navigation includes the intermediate viewpoints during travel. The immediate nature of discrete selection is often compared to teleportation or moving with “infinite velocity” [18].

While this form of viewpoint control is highly efficient, the lack of intermediate details gives rise to criticism. Without a corresponding natural experience, viewers who have been instantly teleported to a new location must take some time to re-establish their bearings [6, 16]. Robertson reports that the intermediate animation “helps the user perceive object constancy, which shifts to the perceptual system work that would otherwise have been required of the cognitive system to reassimilate the view after it had changed” [101]. Similar results are found in the surveillance literature where traditional interfaces require the operator to switch between multiple video streams projected on a small number of displays. Selecting the appropriate stream often requires spatial processing from the operator to relate a specific camera with the desired view [97]. This demand is exacerbated when the remote cameras are mounted on pan-tilt units, requiring the operator to visually interpret the orientation in which the camera was last used.

Recognizing that “there is no set of [interaction] techniques that will maximize performance for all applications and domains”[15], we observe that each method described above offers its own strength. Tan argues that the viewer should be able to toggle between various metaphors based on the particular navigation subtask that is being executed, for example steering for searching toggled with orbiting for objection examination [119]. Others specifically argue against this kind of mixed-metaphor interaction, citing the need for the user to preserve state and other problems that are common to modal interfaces [51]

## 2.3. Cognitive Principles of Viewpoint Interaction

The primary benefit of viewpoint manipulation is to offer alternative varied perspectives, facilitating the understanding of a display. The previous section outlined the mechanics of physical interaction – how is the viewpoint controlled. This section addresses the internal processes of viewpoint interaction. Understanding the cognitive principles related to intake and processing of 3D displays provides insight to motivation behind the physical interaction. Figure 10 gives an overview of the major components that are considered. Perception of the environment details sensory acquisition from the display. Spatial knowledge captures the internal representation of the environment and Awareness deals with attempts to resolve disparities between the internal and external representations.

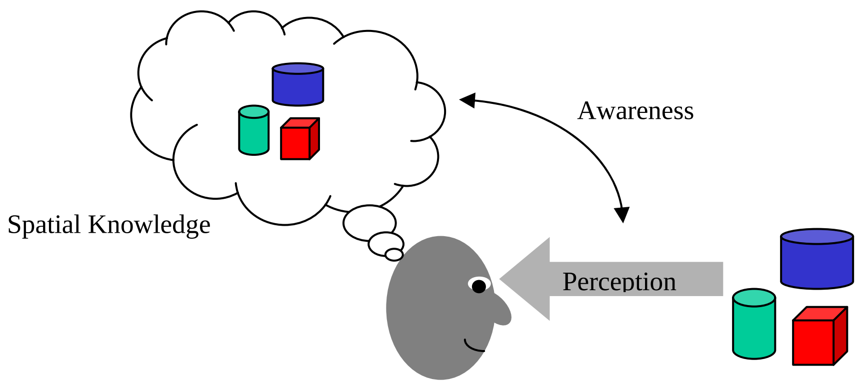



**Figure 10 Cognitive Components of Viewpoint Interaction**

### 2.3.1. Perception of 3D Environment

"No matter how intelligent the choice of information, no matter how ingenious the encoding of the information, and no matter how technologically impressive the production, a graph is a failure if the visual decoding fails"[32]. Manipulating the viewpoint in visualizations leverages certain skills that have evolved to help people navigate in the real world. However, in the real world, other sensorimotor systems supplement information received through the visual channel. While Virtual Reality Systems may try to engage these other senses, this objective does not generalize to the broader field of visualization[3]. In addition to the impoverished sensory system, the visual system is also generally constrained by a display that is relatively small and notably flat. For these reasons, it is important to investigate how the structure of the environment is perceived by the observer.

Given that the majority of visualizations utilize 2D media, effectively communicating 3D representations relies on perceptual trickery. The eye must be instantaneously convinced that the flat display contains depth information. There are a number of depth cues that can be added to visual models, or inferred from sampled environments that assist with this task. These cues can be categorized as monocular static (pictorial), monocular dynamic (moving pictures) and binocular [129].

Static Monocular cues is also known as “pictorial” cues because they can be represented in static pictures, as demonstrated in Figure 11. Linear perspective is the most well known of these cues, perfected in the art community by Filippo Brunelleschi during the Italian Renaissance. At the heart of linear perspective is the fact that distant objects occupy a smaller portion of the retina. At a certain point, a distance is reached where the retinal image becomes so miniscule that it perceptually vanishes. This vanishing point leads to two important properties regarding the perception of depth in images: Convergence – parallel lines will appear to meet in the distance and Relative Size – if two objects considered to be of equal size, but one appears larger, it will be perceived as closer to the viewer. Another depth cue, which could be construed as an extension of the relative size property of linear perspective, is texture gradient. Textured surfaces, viewed from an oblique angle, will show a change in texture density and size [136]. Finer, less articulated textures indicate more distant regions on the surface. A third pictorial cue is known as interposition. This cue shows that nearby objects will occlude distant objects. Finally, lighting

[3] It may be the case that vision and some combination of haptics, audition, proprioception or vestibular engagement may coalesce to an ideal artificial information system. However, the scope of this paper is limited to what can be achieved strictly through the visual channel.

conditions and the resulting shading and shadows also provide information about the depth of the display. The brain seems to be hardwired to assume a single light source coming from above. This explains why the bumps in Figure 11d become hollows if the picture is turned upside down [129].

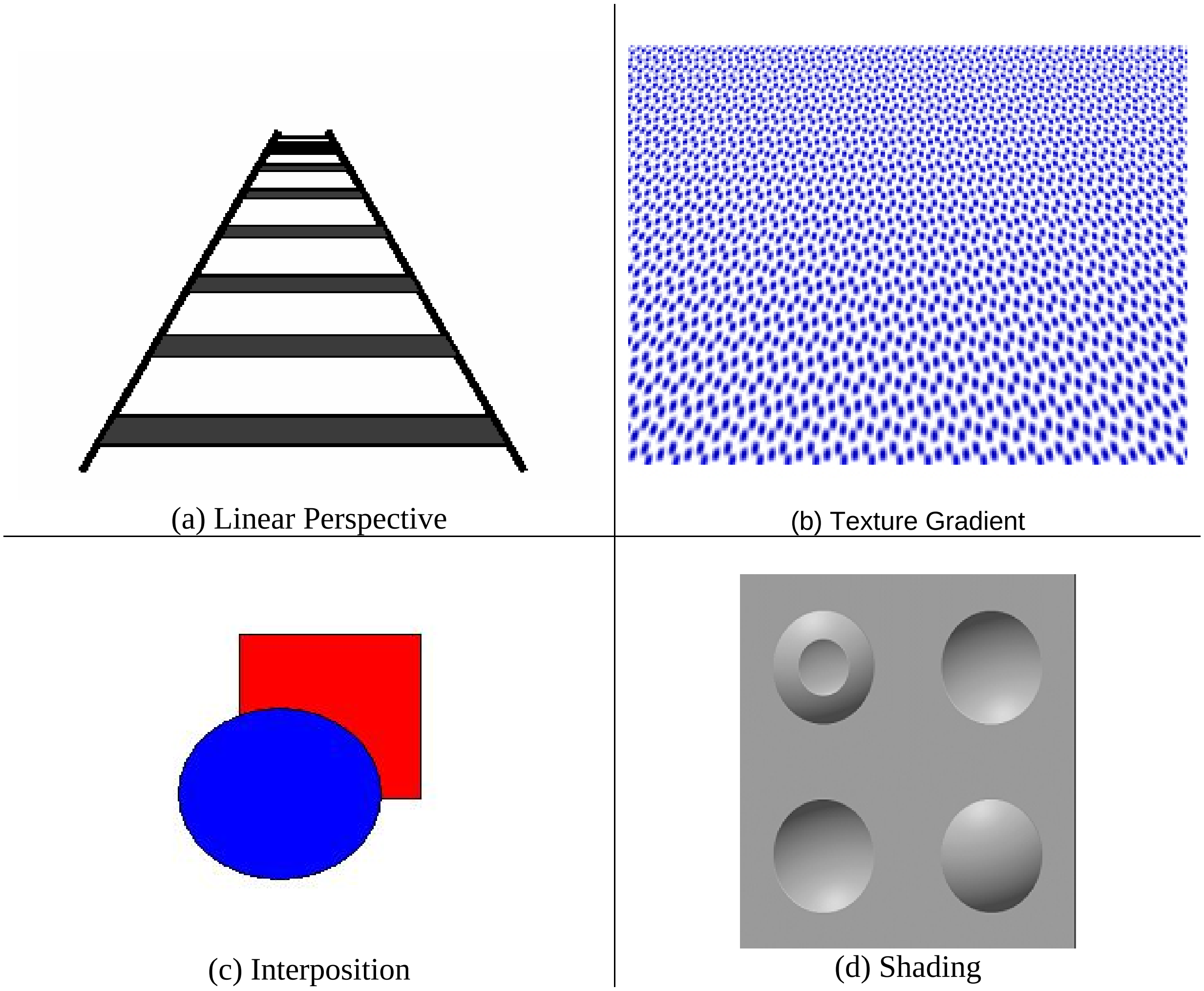



**Figure 11: Static Monocular Depth Cues**

Monocular Dynamic cues, detailed in Figure 12, describe the additional depth information that is provided when motion is added to the display. Motion parallax has certainly been experienced by anyone who has looked out the window of a moving car or train. Distant objects seem to move very slowly relative to the viewer, while nearby objects whiz past at a much faster rate. A second dynamic cue is optic flow. This describes the sense of expansion that one feels when moving forward in the environment (or conversely the sense of contraction when retreating). By paying attention to the optic flow vectors that objects travel along, it is possible to make accurate judgments as to the relative location in the space. Finally, the kinetic depth effect allows a viewer to intuit depth information from a 2D projection of a rotating rigid 3D object [129].

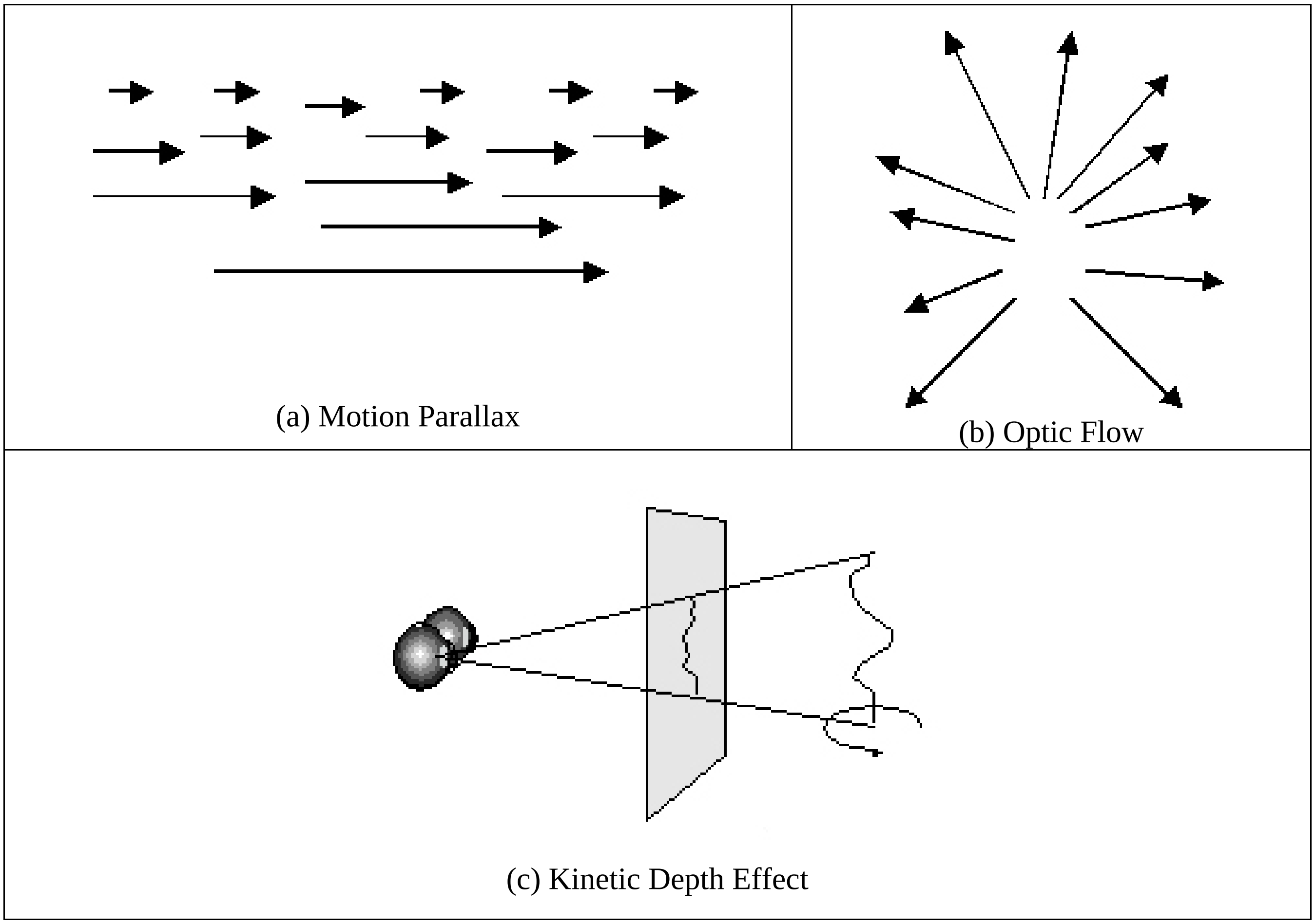



**Figure 12: Monocular Dynamic Depth Cues [129]**

Binocular cues arise from the brain's attempts to resolve the fact that each eye provides a slightly different image of the environment. Adult humans have a span of about six to seven centimeters between their eyes. This is far enough to provide distinctly different viewpoints, capable of conveying extremely reliable depth information for objects within 30 meters [35]. Unfortunately, providing binocular cues in visualization environments often requires specialized equipment [83]. Since stereo vision is based on each eye receiving a distinct image, stereo displays require two images and a mechanism for directing each image to the appropriate eye. This is most frequently accomplished using either a Head-Mounted Display, which projects separate images directly in to each eye, or simultaneously presenting both images and selectively filtering the images with polarized glasses.

Another major factor in the perception of 3D displays is the Field of View (FOV) afforded by the camera. The FOV is typically defined as the angle subtended from the left and right sides of the viewing frustum [36], and effectively defines the width of the view plane. Conventional design recommendations suggest that the FOV should be set to as wide as possible. Narrow FOVs are often cited as problematic, interfering with the ability to integrate information from the environment. [43].

However, there is an important caveat to this guideline. While the human FOV across both eyes spans nearly 200°, the display often occupies only a small portion. For example, depending on the viewing distance, an 18 inch monitor covers only 30° to 40° of our field of view. This means that two FOV measurements should be considered: the Display FOV (DFOV) and the Geometric FOV (GFOV) [42]. The DFOV is relatively constant, determined by the relationship between the viewer and the edges of the display. The GFOV, on the other hand, is an variable setting of the virtual camera. As the GFOV grows, the resulting image will appear smaller. Conversely, the display will be magnified as the GFOV shrinks. Draper demonstrates that as the ratio between these two angles diverges from 1:1, there is greater likelihood of misperception and simulator sickness [42]. This altered sense of scale can also impact the perception of velocity through the environment. Special displays, such as CAVES [34], are designed to give the viewer an immersive feeling by expanding the DFOV. However, desktop systems should probably adopt FOVs appropriate to their size.

### 2.3.2. Spatial Knowledge

The internal representation of spatial knowledge is most often referred to as a *cognitive map;* a term coined in 1948 by Edward Tolman to describe the learning of rats in mazes [113]. Like their external counterparts, cognitive maps are presumed to allow people to internally identify the presence or absence of key landmarks, infer connectivity, distance and direction relationships, and generally organize spatial information. However, Tversky observes that spatial information is rarely stored as a coherent whole that is available for mental inspection. [123]. She argues that a more appropriate metaphor would be a *cognitive collage,* a collection of thematic overlays representing multiple perspectives and levels of specificity.

Acknowledging problems with the terminology, Golledge advocates the term cognitive map to mean the "deliberate and motivated encoding of environmental information so that it can be used to determine where one is at any moment, where specific encoded objects are in surrounding space, how to get from one place to another, or how to communicate spatial knowledge to others" [54]. Regardless of the terminology used to describe this internal representation, it is this conceptual model of the environment that drives the interaction with the external representation of the environment. It is therefore critical to understand the nature of the information being stored and how it is acquired.

Siegel and White's Landmark-Route-Survey model provides a pervasive, generally-accepted taxonomy for the types of spatial knowledge [111] (summarized in Table 5). In this theory, l*andmark knowledge* develops early as significant features are extracted from the environment and remembered. This type of knowledge is often compared to looking at a series of photographs, as the embedded information is static, orientation specific and independent of other landmarks [38]. What constitutes a landmark is somewhat ambiguous as significance may be determined by personal relevance or experience. However, a common characteristic of landmarks is that they are often selected based on visually distinct features that are capable of attracting attention [54]. Serving as a declarative reference point, landmarks are ultimately the key organizing feature of the mental representation.

The second type of spatial knowledge, *route knowledge*, develops when specific connections are formed between the landmarks, allowing the individual to traverse the environment from one landmark to the next in a systematic manner. While these routes can be concatenated to form chains of connectivity, the ability to find a novel, optimal return route (homing) is still not present at this stage; comprehension is limited to memory of prior traversals [134].

Finally, *survey knowledge* is distinguished by the ability to understand the global configuration of the environment. This stage is characterized by the ability to understand the direction and distance between landmarks independent of a specific route. A mental "bird's eye" view of the environment can allow people to develop short-cuts and successfully negotiate fresh routes on-the-fly [38].

**Table 5: Landmark-Route-Survey Model of Spatial Knowledge**

| | Landmark | Route | Survey |
|---|---|---|---|
| **Information Type** | Categorical, Declarative | Topological Procedural | Metric Spatial |
| **Affords:** | Recognition, Presence or Absence of Features | Rote Traversal, Networking | Distance and Directional Estimates, Negotiated Shortcuts |

Although Landmark-Route-Survey knowledge was initially thought to be acquired incrementally in strictly sequential stages, this is not necessarily the case. This misconception is partially based on the assumption that spatial knowledge is constructed by integrating egocentric viewpoints [95]. Factoring in alternative viewpoints, such as looking at a map, it has been determined that survey knowledge can be more directly acquired, bypassing the need for an incremental build up [39, 63]. However, the precipitation of survey knowledge does not mean that studying maps is the best way to learn spatial information. Thorndyke and Hayes-Roth found that map learning produces orientation dependent survey knowledge, and people were prone to errors when their perspective was not aligned with the orientation of the map [121]. Comparatively, survey knowledge that developed from exposure to the environment was not subject to such errors.

Other conditions also factor into the development of survey knowledge, including the scope of the environment and the quality of the exposure. For example, survey knowledge of regions that can be inspected from a small range of viewpoint movements develops more quickly than for distributed regions that must be understood by large traversals [33]. This finding also applies to clearly defined regions within a larger space. For instance, configuration knowledge of objects contained within the same room can be almost instantaneous, while the relationship between objects in different rooms develops is learned at a slower rate. This phenomenon is thus known as the “Room effect”.

Motion is another critical factor in developing spatial knowledge. Observers who view a scene from a single position develop orientation specific representations, regardless of whether the frame of reference is an exocentric map view or an egocentric, first-person perspective [3]. However, motion through the scene allows observers to benefit from the multiple perspectives, making it easier to adapt to an unfamiliar orientation.

While the addition of motion enhances spatial knowledge, it may not be enough to simply move a user through an environment as if on a tour bus; self-controlled users demonstrate better spatial knowledge than passive observers [99]. Self-controlled viewers not only remember what they see, but they also remember the actions they took to get there. Moreover, active viewers receive information on-demand. As such, the information they receive is based on their particular needs, allowing them to examine a portion of the environment until it is understood.

Despite our best efforts, numerous systematic errors are commonly associated with human spatial knowledge acquisition [54]. Moreover, the time that it takes to develop an internal representation means that spatial knowledge is often incomplete. While acquiring a lasting, complete spatial knowledge may not be the primary objective of the visualization, if the viewpoint is to be manipulated extensively, it should be an important secondary goal. Recovering from disorientation or otherwise desiring to revisit previously viewed regions requires some level of spatial knowledge.

### 2.3.3. Awareness

For successful navigation, it is not enough to have detailed survey knowledge of the environment. Nor is it sufficient to be able to accurately perceive the environment. Awareness – synchronization between the visual stimuli with the internal representation – is required for the viewer to fully comprehend the display, make decisions about future movements and maintain orientation while exploring

In theory, by navigating through the environment, spatial knowledge develops and errors, ambiguity, and uncertainty about the environment fade. However, this can only really occur if viewers are actively aware of their location and orientation while exploring. This process is sometimes referred to as “spatial updating”. Golledge describes two strategies for spatial updating: path-integration and piloting [80]. Path-integration allows the viewer to intuit their location based on cumulative movements, relative to a starting point. This technique, also referred to as “dead-reckoning”, permits exploration without any spatial knowledge, but it is subject to accumulated errors, and often needs to be corrected by finding a known location. Since path integration relies on viewers understanding their movements, this approach is particularly susceptible to perceptual errors stemming from optic flow and field of view, described previously. Piloting affords a more precise awareness in the environment by allowing viewers to calculate their position relative to external landmarks. However, if the landmarks are out of the sensory field, piloting presupposes the existence of an accurate internal or external map to provide the necessary context.

Endsley offers a formal model for characterizing awareness. She defines Situation Awareness (SA) as: “The perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future” [47]. This definition can be decomposed into a three-tier model of the stages of awareness. Level 1 SA captures the perception of elements in the environment. This encompasses identifying the “status, attributes and dynamics” salient information in the environment [47]. The second level of SA combines individual elements with each other and with the viewer’s knowledge to form patterns, yielding an understanding of the significance, related to the viewer’s objectives. Finally, the third level of SA integrates the viewer’s comprehension of the current state, with the consequences of future alternatives. This level of awareness empowers viewers to anticipate the results of interaction, which allows them to make informed decisions and project future states.

# 3. Design Considerations for Assisted Viewpoint Interaction

> What information consumes is rather obvious: it consumes the attention of its recipients. Hence a wealth of information creates a poverty of attention, and a need to allocate that attention efficiently among the overabundance of information sources that might consume it.
>
> ~Herb Simon [126]

Describing the physical interactions and the cognitive challenges required to control the viewpoint, it becomes apparent that a high degree of attention is consumed. To make viewpoint manipulation more successful, techniques for assisting the navigation task are beginning to emerge. Such methods may seek to ease cognitive load or facilitate the interaction required to manipulate the viewpoint. They may increase the efficiency or effectiveness with which the environment is explored, by promoting awareness. They may enable the viewers to acquire viewpoints that allow them to understand an artifact or relationships between artifacts that they might have otherwise overlooked, increasing the transfer of salient information. To date, various methods of viewpoint assistance have been proposed largely on the basis of individual, ad-hoc solutions, without an eye toward broader objectives. This section provides a unifying framework of assisted viewpoint interaction that can be used to inform the design of future systems.

## 3.1. Types of Assistance

Human operators are capable adjusting the viewpoint by simultaneously manipulating multiple degrees of freedom. Even with training and sophisticated strategies, there is the chance that the effort applied to manipulating the viewpoint will distract from viewing the environment. Studies of how people interact with 6 Degree-of-Freedom devices reveal that there is a division between interaction with translation and orientation controls [63, 83]. People tend to issue clusters of commands, toggling back and forth between sequences of translations and sequences of rotations. This natural boundary suggests a division of labor between manual and automated viewpoint control, yielding two paradigms: Guided Positioning systems and Guided Orientation systems.

### 3.1.1. Guided Positioning

In a Guided Positioning system, assistance is provided with moving the viewpoint through the environment. There are multiple ways to establish the route, depending on the amount of environmental knowledge afforded to the system. At the most basic level, the route may be a pre-programmed sequence of steps through the environment, offering a generic tour. Generalizing this approach, the viewer may be able to specify a set of interests. Based on this input and other constraints such as obstacles in the environment, it is possible to generate a more personalized tour [45]. When the system has very little foreknowledge of the environment, it will likely adopt a naïve search strategy that systematically moves the viewpoint through the environment. Examples include the lawnmower method, which moves along narrow, adjacent strips, radial search, where exploration progresses in increasing concentric circles or contour following [40].

Motion planning systems can be informed by heuristics for cinematography [95]. Specifically:

- Attempts should be made to not only avoid obstacles but pass them by a wide margin. Near-collisions result in a large part of the screen being occupied by the obstacle
- When executing sharp turns, the speed should be slowed to prevent objects from passing too quickly through the field of view.
- “Dull” shots can be avoided by moving the camera at an acceptably rapid speed.

### 3.1.2. Guided Orientation

Guided Orientation systems offer assistance with rotating the camera to acquire interesting views. In teleoperation, poor positioning can clearly jeopardize the safety and success of the mission, however, establishing proper orientation is just as critical to successful viewpoint interaction. Murphy found perceptual ambiguities rather than navigation, locomotion, or manipulation to be the primary problem in teleoperating search and rescue robots at the World Trade Center [93]. Bajscy outlines two important tasks must be supported to effectively implement guided orientation: shifting and holding [5]. Gaze shifting involves transitioning the focus of the viewpoint from one point of interest in the environment to another, while gaze holding describes the activity of keeping an interest point in focus despite viewpoint movement or other environmental changes. Guided orientation may also be used to provide cues to how the viewpoint is moving. Interactions with environmental factors, such as staircases or corners instinctively cause certain adjustments to our viewpoint. By predictively panning the camera when nearing a turn or tilting when approaching a staircase, a more natural interaction can be achieved [75, 95].

As with motion planning, there have been numerous approaches developed for how to achieve guided orientation. One method binds the ideal orientation to the location in the environment [60]. As the viewer moves through the environment, the position is used to index an array of ideal gaze vectors and automatically align the camera in the appropriate direction. While this approach can be effective for directing attention in static environments, it would not work if the system needed to track a moving object. Alternatively, elements of the environment can be assigned “attraction values”. Aggregating the forces of nearby objects generates a potential field which can be used to manipulate the orientation of the camera [10]. A moving object with a high attraction value could be tracked much as a nearby magnet alters the behavior of a compass needle.

### 3.2. Generating Recommendations

Implicit to the idea of assisted viewing is that the recommender has some knowledge of the viewers’ information needs and how they relate to the environment. Burke classifies several techniques that commonly used for converting this knowledge into recommendations [23]. Collaborative recommendation aggregates ratings or usage profiles and attempts to recognize common trends that occur within subgroups of the population. Recommendations are then made to suggest information that was deemed useful by similar users. In the case of viewing recommendations, the system may track particular paths through and environment, commonly used vistas or viewing angles. Content-based recommendation systems match information needs to a feature list from the environment. For instance, when inspecting a medical visualization, viewers may benefit from having their attention drawn to high-risk regions determined by the presence of certain chemicals or compounds. Knowledge-based recommendations offer a higher level of advice. This class of assistance can bridge the gap between viewers' information needs, e.g. driving a remote vehicle, and the specifics of the environment, e.g. an obstacle in the vehicle's path.

In order to generate effective recommendations, the system must also be able to extract certain semantic knowledge from the data source. The degree to which the system "understands" the data can be expressed as a continuum along two dimensions: granularity, and descriptiveness. The resolution at which recommendations can be applied is limited by the data's semantic *granularity*. A coarse model might have information that explains that it represents a specific building. Alternatively, a fine-grained approach to the same model might provide deeper information about the floor plan, or even include transient features such as furniture. Thus, highly granular data sets can be divided into meaningful segments that provide the ability to discriminate knowledge within portions of a visual display. The second dimension, *descriptiveness,* captures the amount of detail that associated with each granule. Meta-data about each feature may provide only a few keywords to summarize a lengthy concept, or it may adhere to a robust ontology that provides minute detail of the object.

The granularity and descriptiveness of a dataset are heavily influenced by the type of visualization that is being viewed. In general, analog objects are continuous in nature making them resistant to decomposition and more difficult to describe. At the same time, digitized objects are granular, but the granules may not be easily mapped to higher-level components for description. For instance, inferring that a set of pixels within a given color range constitute a straight line of a given color and length is difficult. That same line, as a component of a CAD drawing already has all of the meta-data attached and the routines to manipulate the underlying structure are relatively easy to write. Recall that in the case of information visualizations, the display is completely constructed based on algorithmic processing of the dataset. Since the attribute lists of the individual data points were integral to the construction of the display, suggesting the desired viewing parameters should be an extension of the process used to construct the display in the first place. At the other end of the spectrum, sampled environments inherently contain no information about the elements of the display. Image processing or other sensor data can be parsed to reconstruct a semantic picture of the world, but this often requires complex processing and comes through the filter of the sensors.

Ultimately, the degree of system knowledge will have a profound effect on the generation of recommendations. If there is sufficient knowledge of the environment, recommendations can be derived or automatically processed. While this offers a great deal of flexibility, embedding knowledge in the viewing environment requires a significant overhead in the form of data structures, processing languages and knowledge bases that are robust enough to support this kind of computation. Alternatively, knowledge can remain in the head of an expert who marks-up visualizations with viewing instructions. Clearly it is not practical to rely solely on a manual approach, however, this tradeoff may be worthwhile in certain training settings with more tightly constrained viewing objectives or real-time scenarios where the cost of machine intelligence is presently prohibitive.

### 3.3. Informing the Viewing Experience

Regardless of how the system knowledge is acquired, numerous methods have been explored for how the recommendations can be instantiated to facilitate the flow of relevant information to the viewer. These techniques can be characterized on a continuum of system engagement. Some interfaces may provide information that the viewer passively assimilates, while others may take a more active role in redirecting the viewpoint. Brusilovsky provides a taxonomy of several popular technologies used in hypertext recommendation systems including annotation, hiding, and sorting [22]. This framework can be abstracted to characterize the methods for informing the viewing experience with 3D visualizations.

### 3.3.1. Annotation and External Cues

As discussed in the previously, there are large parallels between wayfinding in natural settings and virtual environments or visualizations. To overcome some of the deficiencies with spatial cognition in natural settings, travelers often rely on external assistance from maps compasses, etc. These assisting devices have extended to navigation in virtual environments as well. External cues can supplement the content information by providing additional navigational information ranging from passive contextual awareness to active guided tours through the environment.

Heads-up displays (HUDs) were invented project operational information onto the forward field of view, allowing operators to acquire contextual information without needing to divert their attention to a separate control panel. In immersive environments, it is common to display heading information is displayed in the form of a compass. This allows the viewer to make judgments about their relative direction of travel. Depth bars, as shown in Figure 13, are another annotation that can help understand depth, especially in naturalistic settings [49].

**Figure 13: Depth Bars [49]**

Providing a map of the environment can also be an effective tool when interacting with immersive visualizations. Darken found that maps act as a supplement to survey knowledge and that using a map while navigating a virtual environment resulted in better search strategies [40].

HUDs can also be used to assist with perceptual enhancement and offering suggestions. Zhai describes a visualization interface where irrelevant features of a display are slightly darkened, which has the effect of making a desired portion of the display brighter and more instantly recognizable in a search task [144]. The physical analog for this kind of display is a spotlight that highlights important information (see Figure 14a). An interface that used a computer-controlled spotlight to detect and fixate on objects of interest in a 3D search has been shown to have a significant reduction in search time, misidentification, and omission of critical information [68]. Moreover, varying the intensity of the spotlight can effectively convey information about the confidence of the system's suggestion [69]. The downside of using a spotlight is that it can only direct attention to elements of the environment that are within the field of view. An alternative approach is to indicate the vector that the spotlight would shine through some external imagery. This could take the form of an "intelligent" compass, or an arrow projected to the floor of the environment (see Figure 14b). Likewise, Satalich designed a system to help guide people through architectural structures that consisted of drawing lines on the floor that connected points of interest [105].

Similar navigation assistance can also be provided in the form of an avatar, embodied agent, or other animated character. Chittaro proposes that using these embodied guides are particularly useful for explorers who are new to an environment and serve multiple purposes [31]. First, the agent can induce the viewer to follow, leading to discovery and awareness of a particular region.

Once there, it can further direct attention to details using natural gesture and elaborate with additional information, perhaps audio or textual descriptions.

The shape assumed by the guide may have a significant impact on the expectations for the assistance that it can provide. It may be desirable for the agent to assume an anthropomorphic shape, as shown in Figure 14c [67]. Since we are to dealing with human guides, we can tap into unspoken cues such as shared gaze directions, projectability and other forms of body language [77]. However, personal guides that are quite good a providing wayfinding assistance may be inadequate for providing other kinds of assistance [125]. Assuming a human form may lead to assumptions that are unsupportable by the current state of technology, including natural language, or simply beyond the scope of the visualization task. For this reason it might be beneficial to adopt an alternative representation that is more appropriate to navigation. Wernert has suggested an implementation where the viewer is walking a dog a guide avatar – by tugging on the leash, the avatar is suggests other routes or directions of interest [137]. Another interesting approach involves "walking products" [30]. Multiple avatars walk through the environment bearing an image of their destination. In order to locate a specific object or class of objects, the viewer need only follow the appropriate avatar.

(a) (b) (c)

**Figure 14: Annotations [67]**

On-screen annotations can be effective with providing awareness cues and guidance information. However, designers should be aware that they also take up valuable space, cluttering the display, causing attention competition [142], and potentially occluding important information [12]. Moreover, powering the animation for complex guides, such as the animated characters, may tax the graphics capabilities of the visualization system and detract from the overall system performance.

### 3.3.2. Restriction

Comparable to Brusilovsky's "hiding", this approach involves restricting the number of viewing options to a limited subset. Irrelevant paths or vistas are concealed, leaving the viewer to choose only from alternatives that are consistent with the task at hand. This can be achieved by either limiting input parameters or by selectively paring the viewing options according to the environmental components.

The most basic way to restrict viewpoint interaction is to limit the degrees of freedom that the viewer manages. Several methods were previously described for restricting input in order to deal with hardware limitations. Reducing the control space is also appropriate depending on the task facing the viewer. Many applications allow the viewer to manipulate 6DOF, even when that is not necessary. According to Bowman, the higher the dimensionality of the path traveled, the more likely the viewer is to forget information seen along the path; limiting the viewer to the ground plane (when appropriate) can significantly reduce the cognitive load [14].

Coupling can be effective, provided that care is taken to preserve the necessary interactions to complete the task. Gaze-directed steering for example can be very intuitive for quickly exploring a scene. However, tasks that require extensive inspection of objects in the environment benefit from independently controlled orientation of the camera [14, 70]. Another example of effective coupling is known as speed-coupled flying [120]. This technique attempts to infer the need for detailed views from the speed with which the viewer travels. If the viewer moves rapidly through the environment, the system assumes they know where they are going and floats the camera into the air and tilts it downward, providing a overview of the environment. As the viewer slows down, the viewpoint gradually descends back to the ground plane, allowing for more detailed inspections of the contents.

Discrete selection of the viewing parameters is another mechanism for restricting input. QuickTime VR is one popular system that offers this kind of interaction [27]. Using this application, viewers can navigate through a space by selecting critical viewing locations from a discrete list. Once the camera location is loaded, the viewer is able to adjust the viewing orientation by panning and tilting the camera according to their wishes.

Restriction does not have to occur along the boundaries of a particular motion variable. For instance, in an object-centered display, it is clearly not beneficial for the viewer to pan the camera 180° away from the object being viewed; there would be nothing to see. This motivates the Virtual Trackball method, where the camera can be moved along a the surface of a sphere and the camera always points toward the center [26]. Even though the viewer is only controlling motion on a surface, they are making use of all three dimensions for translation purposes. For immersive environments, Hanson extends this concept to general constraint surfaces [59]. Figure 15 shows three example constraint surfaces that follow a terrain to varying levels of specificity.

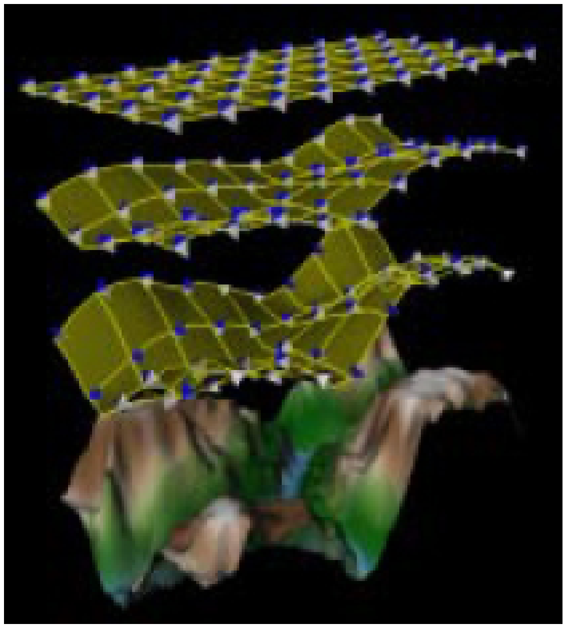


**Figure 15: Constraint Surface**

The real insight behind this concept is that the camera does not necessarily have to explicitly follow the topology of the environment. Instead the visualization designer can use the constraint surface to enhance certain views and limit access irrelevant regions. For example, warping and folding the constraint surface can achieve various effects, including vista points, multiple-coverings and spiral elevators [60]. A constraint surface may not even cover the entire environment, but instead restrict motion to a virtual sidewalk through the scene. Gaylean proposes the "River analogy" in which movement of the camera is constrained to an imaginary river flowing through the environment [53]. Like operating a boat, the viewer still determines position and rate of speed, but the contour of the river restricts the extent of travel. These techniques may be used to increase the likelihood that the camera will pass certain features of the environment and can promote a sequence in which they are viewed.

One drawback to this approach is that it may be difficult for the viewer to understand the effects of the constraint surface. In the case of the virtual trackball, the viewer can quickly learn to rely on the regularity of the sphere, but when following a virtual sidewalk, the boundaries, and thus the next step may not be as apparent. Moreover, since constraint surfaces are not always aligned with environmental topology, viewers may become confused when the constrained path is at odds with visual cues used to determine a path through the environment.

Taking the restriction paradigm to the extreme, direct guidance or automation is achieved when a strictly linear order is established through the navigation space. Instead of giving the user options to navigate, the "best" choice is selected by the system. In terms of hypermedia systems, pages are presented with only one link, usually labeled as "next". For 3D environments, direct guidance is accomplished by presenting a predetermined sequence of viewpoints to the viewer. This is equivalent to viewing a movie where the director has determined both the camera position and orientation. Numerous systems have been developed for converting high-level tasks ("Show me all of the objects with property X") into dynamically generated viewpoint sequences [7, 45, 57, 61].

As all kinds of restriction, there is a serious danger with using direct guidance: inaccurate restrictions. If the viewer is restricted from acquiring a viewpoint that (they think) is necessary, it will quickly lead to viewer frustration.

### 3.3.3. Prioritization

In Brusilovsky's taxonomy for hypertext systems, "sorting" presents a list of navigation links arranged by relevance. Clearly kind of restructuring cannot be applied to all 3D displays. One of the hallmarks of scientific visualizations is the attempt to capitalize on consistency with the real world. Arbitrarily reorganizing the navigation options could seriously compromise the fidelity of a model. However, the intent of sorting as a recommendation technique is to facilitate the selection of a particular option while still presenting the alternatives; it is easier to select an option from the beginning of the list. In this spirit, it is possible to accomplish "sorting" in a visual display by making easier for a viewer to acquire a particular view of the data. There are several mechanisms that can prioritize certain views while maintaining the structural integrity.

One way to implement prioritization is through the systems response to input. Mackinlay's POI navigation adjusts the viewer's motion speed logarithmically in relation to the distance from an object of interest [82]. Thus if the viewpoint is far away from an interesting feature, only a small movement is required to bring the viewer closer to the goal. Conversely, if the viewer is nearby relevant artifacts, it takes more effort to move away from it. A similar type of prioritization could be realized with the emergence of haptic input devices, which could hypothetically offer tactile feedback and resistance to inappropriate commands offered by the viewer- encouraging them to take the path of least resistance.

Prioritization can also be achieved by introducing automation that can be overridden by the viewer. In this sense, high priority views will be displayed by default, but with some extra effort, the viewer still has access to views that are deemed less important by the system. For example, Beckhaus describes a guided tour system that can be interrupted to allow viewers to explore an area of local interest. When finished, the guidance is capable of resuming from wherever the viewer left off [10]. However, if switching between modes is strictly user-initiated, the viewer effectively needs to know when assistance is required. Unfortunately, this knowledge is often lacking, especially in novice viewers. Merrill suggests that "Learners should be given control, but if they do not make good use of it, the program should intervene and lead the student through instruction" [88].

Establishing a protocol for seizing or abdicating control is particularly relevant to viewpoint manipulation. The viewpoint is often a single shared resource; when one actor manipulates it out of range, it may severely impair progress from the other players in the mixed-initiative environment.

Consider the frustration that would occur if the viewpoint were automatically moved to the next region before the viewer has had a chance to finish inspecting an element of interest. Perhaps more distressing would be the condition where the viewer is ready to move on, but he cannot see where he is going because the gaze is still automatically fixated on an object that has already been reviewed. By some measures, it may still be possible to "succeed" in these conditions, however, failure to address these issues could make an interface design untenable for practical use. Multiple strategies exist for negotiating control, however they can be classified into two major groups: Conditional and Coordinated controls.

Under conditional control paradigms, viewpoint control is determined by certain environmental conditions; when these conditions are met, control of the camera is transferred from the viewer to the autonomous agent and vice versa. One of the most common instances of conditional control is to define "regions of interest". The Attentive Camera, for example, dictates that when the viewer enters the area surrounding an object of interest, the viewpoint automatically pans to center the target object in the field of view [68]. Likewise, StyleCam, provides animated transitions between regions intended for viewer-controlled inspections [24]. Explicit rules may also be provided based on other environmental and operational conditions. Breummer advocates a "safe-mode" for teleoperated robots, where the robot is equipped with a set of rules that allow it to "veto dangerous human commands to avoid running into obstacles or tipping itself over" [21]. The conditions can even be as broad as adopting a policy of non-interference – automation would only occur when the viewer is not issuing any direct commands [137].

As the conditions become more sophisticated, a higher degree of coordination exists between the viewer and the automation. Highly coordinated controls may go as far as employing a mixed-initiative interaction model that assumes that the viewer and the autonomous agent are working together to achieve a goal and engage in a dialogue to “exchange ideas and resolve differences [50]. Instead of a master-slave relationship, where an operator delegates tasks to a surrogate, the burden of determining when certain actions should be performed is shared. Execution of subtasks are “opportunistically negotiated”, depending on who is best equipped to meet an objective [62]. Under this paradigm, the system may need to ask for control of the camera, in order to disambiguate views, or show off important findings, but the operator may deny the request, pending complete a critical task.

Moving toward the principles of mixed-initiative automation to the viewpoint interaction problem could provide the necessary support to ensure that meaningful views are acquired, while still leveraging the human operator’ perceptual discrimination, cognitive flexibility, spontaneity, and ingenuity. Previous mixed-initiative robotic systems have emphasized positioning operations and path-planning for the robot [49], however, providing prioritized recommendations to viewing orientation is yet relatively unexplored.

# 4. Assisted Viewpoint Control System Development

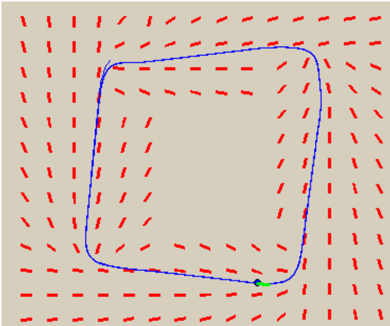
This chapter provides describes the implementation of a specific system developed for assisted viewpoint control. While the base technique could support many of the design considerations described in the previous chapter, evaluation of this technique focuses on interaction with prioritized viewing orientation recommendations.

## 4.1. Attentive Navigation Implementation Details

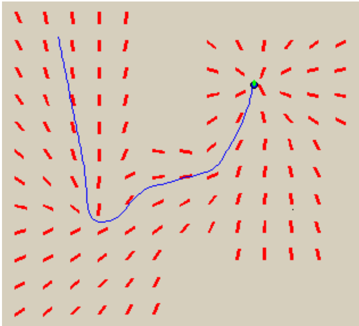
*Attentive navigation* is a specific instance of an assisted viewpoint interaction system that generates viewing recommendations based on the user's context within the environment. This technique is an implementation of a method proposed by Hanson and Wernert in an attempt to facilitate 3D navigation using 2D controllers [59]. Attentive navigation divides the navigation space into two components: 1) the *constraint space*, which defines the range of positional values that can be reached by the input device, and 2) the *camera model field* (CMF) which describes a set of ideal viewing parameters for each point in the constraint space. As the viewpoint moves through the constraint space, the system determines the corresponding values in the CMF and presents them to the viewer.

Practically speaking, the CMF is not explicitly represented; as the number of points that comprise the constraint surface can be arbitrarily large. Instead, the CMF samples control points at a fixed resolution and then interpolates the viewing parameters using a bi-cubic Catmull-Rom spline. This method has the desirable property that the interpolated values will include the control points and ensures smooth, continuous transitions between the viewing parameters [47]. Thus, attentive navigation effectively describes a recommendation engine capable of guiding the viewer to ideal viewpoints based on the context of their exploration. Depending on the semantics of the CMF, this technique can be used to implement guided positioning or guided orientation.

To use attentive navigation to assist with positioning, the CMF stores a motion vector indicating the ideal next position for the viewpoint. At a broad level, the constraint space provides arbitrary boundaries, allowing the creation of virtual pathways through the environment. The cumulative effect of the motion vectors in the CMF can further constrain the movement through the display, defining explicit sequences, resulting in either cyclic paths or terminal points, as shown in Figure 16.



(a)



(b)

**Figure 16**: **Guided Positioning to Generate (a) Cyclic Path (b) Terminal Movie**

To implement guided orientation, the CMF can store a viewing vector, specifying the recommended camera orientation for every location in the viewing environment. Using this approach, it is possible to focus the viewpoint on elements of the scene that build knowledge while impeding perspectives that detract from learning. Even if the viewer is controlling the position of the viewpoint, the system can force the viewers to fixate on important features as they move past, as demonstrated in Figure 17. When a viewer walks along the path defined by the dark gray arrows, the camera orientation is redirected by the CMF vectors (light gray) to fixate on the star.

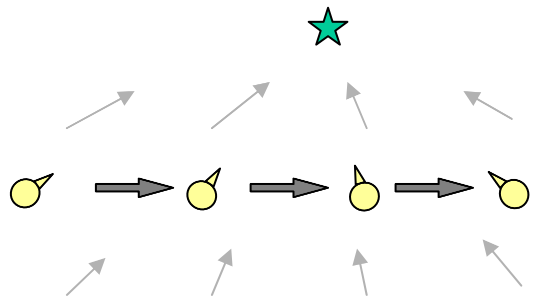


**Figure 17: Gaze Fixation**

As discussed in Hughes et. al.[67], attentive navigation generates display independent recommendations and thus can be used to power the mechanisms for viewpoint recommendation discussed in section 3.3: annotation, restriction and prioritization.

## 4.2. Conditional Assistance for Viewpoint Interaction

A collection of experiments was designed to assess the effectiveness of guided orientation using prioritization powered by attentive navigation. The initial implementation employed a conditional strategy for determining when to make recommendations. The viewer could manipulate the camera location, while the system automatically redirected the viewer's gaze to align with the ideal viewing information for the current position. Unlike a rigorous restriction approach, the alternative viewing options were still accessible; the viewers could adjust the viewing orientation manually whenever they were stationary.

Figure 18 shows a simple 2-dimensional environment map with ideal gaze-direction vectors overlaid. This example shows two points of interest – the lower left and the right center – that will be recommended if the viewer is nearby. It should be noted that this example only shows 2-dimensional vectors, effectively adjusting the yaw of the viewpoint. It is possible to extend attentive navigation to manipulate any number of viewing parameters, including pitch, roll and even vertical position.

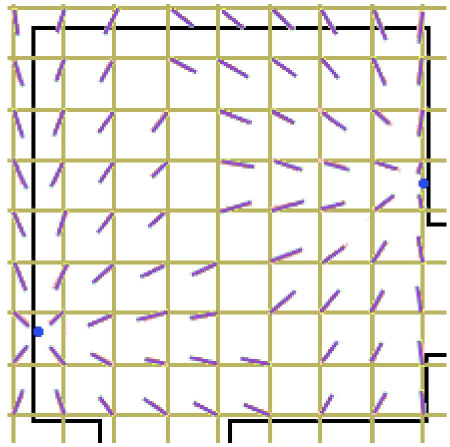


**Figure 18: 2-Dimensional Ideal Gaze Vectors**

These experiments present a range of comparisons that reflect various objectives and interests in the topic of assisted viewpoint interaction. The initial experiment contrasted the effects of using a mouse with automated viewing parameters with the robust control afforded by the use of a 6DOF controller. The second experiment studied the effect of attentive navigation on survey knowledge, and the third experiment examined how attentive techniques can aid with search tasks. Specifically, these experiments are aimed at assessing the validity of the following hypotheses:

H1: Meaningful interaction with 3D environments can be reduced from 6DOF using attentive navigation to assist with viewpoint interaction

H2: Participants exploring with attentive navigation input will develop a better landmark knowledge that with unassisted navigation.

H3: Development of survey knowledge can be enhanced with the use of attentive navigation compared to unassisted navigation

H4: Search tasks are more thoroughly executed with the use of attentive navigation compared to unassisted navigation

In total, these three experiments should provide a more complete picture of the effects and potential uses of assisted viewpoint interaction.

### 4.2.1. Experiment 1: Landmark Recognition

The first experiment was formulated as an attempt to show the viability of using a mouse with attentive navigation as a 3D navigation tool. Previous studies have suggested that a mouse is undesirable for 3D navigation compared to special 6DOF input devices, such as a spaceball or a floating mouse. Although it is possible to simulate 6DOF with a mouse, to do so requires an artificial mapping, resulting in an extra cognitive burden on the user and slower performance [8]. It is hypothesized that attentive navigation can effectively constrain the viewpoint to meaningful orientations, while still allowing operators the freedom to explore on their own terms, resulting in better landmark knowledge development. Hanson's technique was initially conceived as a way to investigate spatial environments with low degree of freedom controllers such as the mouse. Successfully providing such a technique would allow proliferation of 3D visualization tools since users do not require special devices (such as a 6DOF controller), and it allows them to leverage their familiarity with their existing controller (the mouse). A direct comparison of these techniques hopes to provide evidence that attentive navigation achieves these goals.

**Method**

*Participants:* Twenty undergraduate students were randomly assigned to navigate with either attentive navigation or unassisted navigation. Each was paid for their involvement in this study. Given a dependence on vision and identification of color objects, it was stipulated that participation required a self-report of normal or corrected-to-normal color vision.

*Stimuli.* All of the experiments in this study were conducted using a 17-inch color display rendering environments generated in C++ using OpenGL libraries [97].

For this experiment, the environment consisted of a large square room divided into even quadrants. The four exterior walls were shaded light green, and tiled with a brick-like pattern. The four interior walls were light gray had the same brick pattern. The floor and the ceiling both had an

8x8 checkerboard pattern of black and gray squares. All eight walls extended from the floor to the ceiling. The environment also contained a variable configuration of objects. Each object was one of three shapes: sphere, pyramid or a four-sided box, and one of nine colors: red, blue, green, yellow, violet, cyan, orange, gray and light green. Between seven and nine objects were added to the to the room ensuring that no color-shape pair was repeated. Figure 4 displays a sample map of the environment to give a sense of density of objects with respect to the size of the environment.

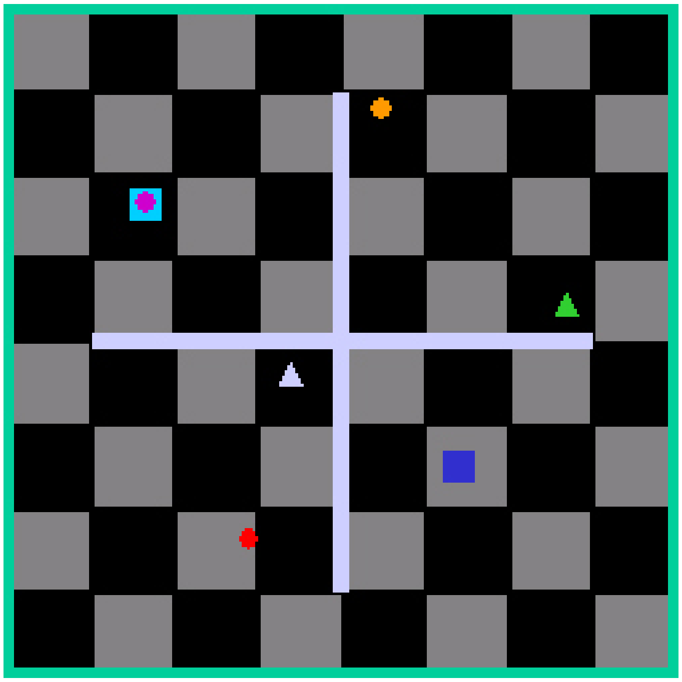


**Figure 19: Sample Environment (Experiment 1)**

The environment was designed to prevent the user from being able to see the contents of an entire quadrant from a single viewpoint. To be confident about the contents of a quadrant, the subject needed to make use of all directions of motion. This was accomplished by the existence of three types of objects. An object was classified as *elevated* if it was positioned such that it only viewable if the viewpoint were adjusted vertically, either through position (+Y) or orientation (+Pitch). An object was *occluded* if it was contained within another object (e.g. a four-sided box). Unless the viewpoint was properly positioned, the viewer would see only the container object. For example, if an open face were on top of a four-sided box, an interior object could only be viewed if the position were elevated (+Y) and the viewer were looking down (-pitch). Finally *plain* objects were defined as any object that was not elevated or occluded. Most of the objects that were encountered were plain, with elevated and occluded objects accounting for two to four objects per environment.

*Apparatus.* Participants exploring with attentive navigation used a standard mouse as the input device, adopting the gaze-directed walking metaphor [17]. Movement was registered by displacement from the initial starting position while the mouse button was depressed. The magnitude of the displacement was translated into a velocity in the VE. Moving the mouse forward/backward resulted in motion in the environment, while right/left motion caused the user to pivot clockwise or counter-clockwise in place. Users were restricted from moving and pivoting simultaneously. Vertical position, as well as pitch and yaw orientation, were influenced by ideal vectors stored in the CMF. The CMF was constructed so that attentive navigation would fixate on the closest object to the viewer. This strategy afforded the opportunity for the viewer's attention to be redirected to each object in the environment.

Unassisted navigation adopted a standard flying metaphor and used a spaceball to register input. Subjects were allowed the following control: Move Forward/Backward, Move Left/Right, Elevate/Descend, Rotate gaze to the Left/Right, and Rotate gaze Up/Down. The velocity was a function of the amount of pressure applied to the Spaceball; more force exerted by the subject resulted in faster movement. Roll transformations were not included for either condition.

*Procedures*. The participants were given a verbal description of the technique, and instruction on how to use the input device to perform certain actions. Additionally, to give a clear mental image of the technique's operations, the experimenter then physically demonstrated how each control movement corresponded to movement and orientation of a person's viewpoint.

Before testing, subjects were placed in a sample environment and directed to familiarize themselves with the control of the navigation technique. They were instructed to train until they indicated that they felt completely comfortable with the operation of the viewpoint manipulation technique and controller. Subjects were required to complete a simple set of motions to verify that they had obtained a baseline of control before proceeding. This verification called for the participant to demonstrate motion all directions followed by travel back to a specific object near the starting position.

After the training was completed, the participants were asked to explore a series of four environments with the task of locating and remembering as many objects as possible. No instruction about the nature or location of the objects was provided to the participants. The exploration session started when the subject pressed the spacebar, and lasted for four minutes. When the time expired, the screen automatically went blank. Immediately following the exploration, participants were given a list of fifteen objects (color-shape pairs). They were asked to identify which of these objects they recalled seeing during their explorations by checking "yes" or "no" next to each entry on the list. The lists were comprised of all the objects that were present in the environment, with the balance assigned randomly from the remaining color-shape pairs. To assist with the recognition, participants could also scroll through a display that rendered each object on the list, providing a visual representation of the color-shape pair.

**Results**

Data were analyzed for the likelihood that an object was overlooked or forgotten for each of the three types of objects: plain, elevated, and occluded. Figure 20 shows that there was no difference between attentive and unassisted navigation for plain objects. However, a single-factor Analysis of Variance (ANOVA) showed that elevated and occluded objects were significantly more likely to be missed using unassisted navigation: $F(1,18) = 16.87$ and $F(1,18) = 20.33$, $p < .01$, respectively.

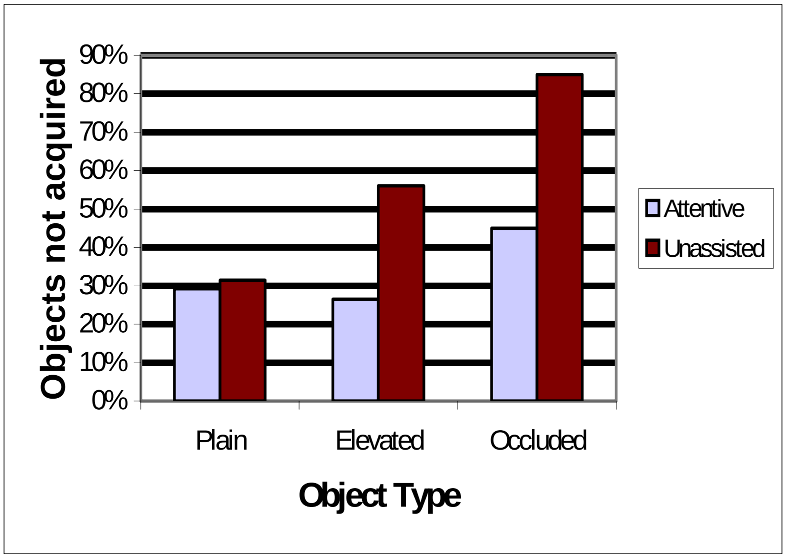



**Figure 20: Errors in Object Identification (Experiment 1)**

In order for the viewer to perform a correct identification, an object must have been in the field of view, and the viewer must have noticed and remembered it. A breakdown of the data into these components provided additional insight to the advantages of attentive navigation.

The display accuracy is the ratio of objects in the room vs. objects that were displayed on the screen. Objects could have been overlooked if the viewpoint was not orientated properly. Using unassisted navigation, this burden rested with the viewer, while the attentive techniques assumed this responsibility. Accordingly, the attentive technique provided a significantly higher display accuracy for elevated objects (+21%, $F(1,18) = 7.05$, $p < .05$) and occluded objects (+35%, $F(1,18) = 10.36$, $p<.01$)

Attentive navigation did an enormous service by getting the objects in the field of view. In some cases this was sufficient. Given that an elevated object was on the screen, there was no difference between attentive and unassisted navigation, $F(1,18) = 1.31$, $p= .27$. in the ability to recall it. However, for occluded objects, attentive navigation provided additional conspicuity; 82% of occluded objects that were displayed were remembered versus only 35% with the unassisted condition ($F(1,18) = 10.18$, $p < .01$).

These results support hypotheses H1 and H2. Attentive navigation limited the navigation to relevant viewpoints, automatically redirecting the gaze to targets while discouraging fruitless searches. It was notable that users of the unassisted condition rarely made use of all the degrees of freedom provided to them; the vertical components of control were often neglected. Generally, users would occasionally remember to search for elevated objects, but often discontinued this practice after lack of success. On the contrary, attentive navigation users could focus on moving the viewpoint through the environment and relied on the automation to indicate when special targets were nearby.

This experiment potentially confounds the benefits of assisted viewpoint interaction with physical operation of the controller. The superior performance of attentive navigation could come from either the automatic gaze redirection or the lack of familiarity with the use of the spaceball. Regardless of this confound, these results show that attentive navigation is an effective control technique that allows users to operate in desktop virtual environments with low degree of freedom controllers.

### 4.2.2. Experiment 2: Survey Knowledge Development

The second experiment was designed to assess the impact of attentive navigation on survey knowledge – the understanding of the configuration of landmarks in the environment. Although this kind of spatial knowledge is facilitated by active, self-controlled navigation, distractions in the environment can be a significant impediment. By using attentive navigation to focus on key landmarks, it is proposed that acquisition of survey knowledge can be hastened.

## Method

*Participants.* Twenty participants, unique to this study, were randomly assigned to navigate with either attentive navigation or an unassisted condition. Each was paid at the conclusion of the study. Given a dependence on vision and identification of color objects, it was stipulated that participation required a self-report of normal or corrected-to-normal color vision.

*Stimuli:* The ability to gain survey knowledge depends somewhat on the topology of the environment. For this experiment, a simple space was constructed that adhered to the following principles:

- A survey of the entire room could not be obtained from a single vantage point. Two large pillars separated the rectangular room into a figure-8 shaped collection of corridors.
- A checkerboard floor pattern was in place to assist the user in estimating distance traveled. This feature also guaranteed that that a visual flow would always be present whenever the user was mobile.
- Different colored walls allowed the user to maintain orientation in a symmetric room.

Each environment contained eighteen objects; three shapes (sphere, cube and cone) by six colors (red, yellow, orange, green, blue, and violet). Unlike the previous experiment, all eighteen combinations were present in every environment. Furthermore, all objects could be considered plain; there were no containers, and all objects rested directly on the floor. A sample map is included in Figure 21.

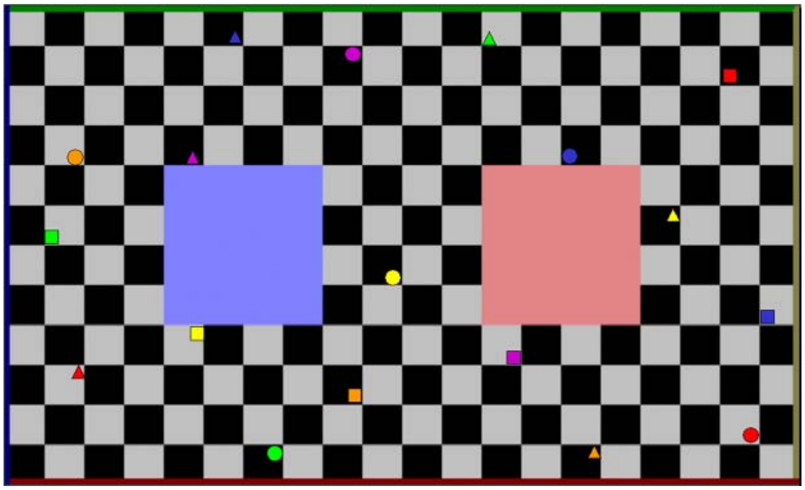


**Figure 21: Sample Environment (Experiment 2)**

*Apparatus.* Since all of the objects were situated on the ground, there was no need for the operator to adjust the vertical components of the viewpoint. Therefore, both groups could navigate using a standard mouse, implementing the gaze-directed walking metaphor described in the previous experiment.

The CMF for the attentive navigation condition was generated in order to manipulate the yaw orientation of the operator's gaze based on proximity to target objects. For this experiment, a subset of the objects in the room was designated as targets of gaze redirection (attended objects). The remainder of objects did not influence navigation (unattended objects). In the first trial, ideal viewing information was provided to focus on the six spheres. In the second trial, the viewpoint was manipulated to fixate in the direction of nearby cubes. Participants were given no a priori information about what attentive navigation would highlight. Although there was nothing inherently important about the Attended objects, this approach can simulate an exploration where a naïve viewer lacks the knowledge to discriminate between relevant and irrelevant features.

*Procedures*. Like the previous experiment, participants were given a verbal description, a physical demonstration, and an opportunity to practice using the assigned technique. Participants were exposed to two timed trials where they were asked to explore the environment with the goal of being able to reconstruct a map. To compensate for the larger environment and to allow adequate time for initial survey knowledge to develop, each trial lasted fifteen minutes. Upon completion of the task, the screen went blank and, participants were given an electronic map framework (boundary lines with the floor pattern) similar to what is shown in Figure 21. Participants were asked to drag-and-drop the eighteen objects from the bottom of the screen to their corresponding positions on the map.

**Results**

The reconstructed maps were analyzed for object misplacement. For each object, the Euclidean distance between actual location and the reported location was calculated. The units of measure used for this analysis is based on the unit distance in the software package used. However, for clarity of this discussion, these measurements, shown in Figure 22, have been normalized with the floor tiles.

There were no differences in the overall error measure between participants using the attentive navigation technique and the unassisted navigation technique. Subjects were able to position an average of approximately eight objects within two tiles of the actual location. However, further analysis reveals a systematic difference in the accuracy with which objects were positioned. Although there was no apparent pattern to the objects well positioned by the unassisted navigators, the objects correctly positioned by viewers navigating with attentive navigation correspond to the objects targeted by the CMF. Figure 22 indicates that users of the attentive navigation technique made significantly fewer placement errors on attended objects than unattended objects ($F(1,18) = 5.24$, $p < .05$). With no distinction between attended and unattended objects presented to the unassisted navigation group during exploration, it is not surprising that there is no difference is observed in map placement ($F(1,18) = 1.38$, $p = .26$). Users of attentive navigation also proved significantly better at placing the attended objects when compared to the unassisted navigation ($F(1,18) = 6.39$, $p < .05$). These results suggest that subtly redirecting attention to a subset of important features can bias the order in which survey knowledge is acquired.

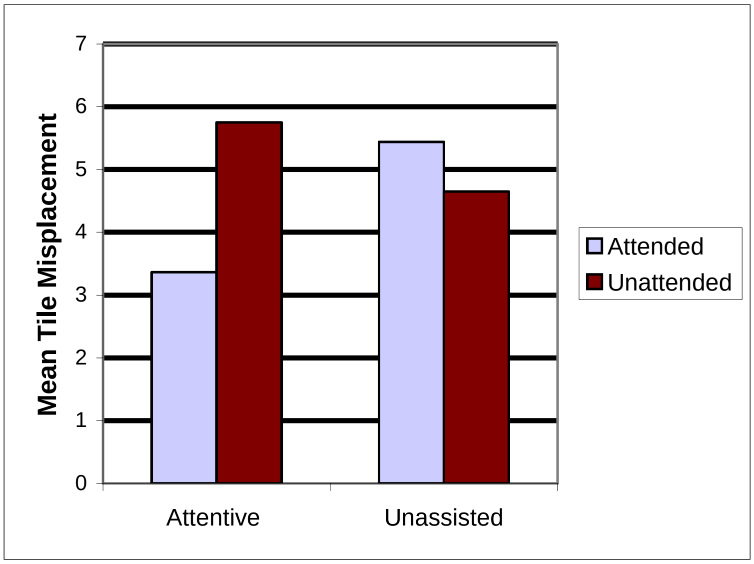



**Figure 22: Object Displacement (Experiment 2)**

It is also worth noting that a partial ceiling effect was observed in the success of attentive navigation. Two of the ten participants using the attentive navigation technique actually placed the unattended objects as well as the attended objects, with a total average error of just over 1 tile per object. Had the environment been more complex, it is expected that the attended score would have stayed low, while the unattended score would increase.

### 4.2.3. Experiment 3: Search Effectiveness

In an environment where fine details need to be processed, it is critical that viewers can quickly hone in on appropriate information. The third experiment was designed to assess the ability of attentive navigation to filter out extraneous details and focus on attributes that are relevant to a given task. By using the attentive techniques in an environment that calls for judgments to be made based on discriminating features, it was hypothesized that this system will enhance the overall effectiveness of a multi-target search task by ensuring that critical features are not overlooked.

**Method**

*Participants.* 26 Participants (13 per condition) were recruited from the University of Pittsburgh community, with most enrolled as undergraduates in the School of Information Science. Upon completion, they were compensated for their involvement in this study. Given a dependence on vision and identification of color objects, it was stipulated that participation required a self-report of normal or corrected-to-normal color vision.

*Stimuli.* All participants were asked to use a joystick to explore an environment with the goal of finding and identifying certain target objects. Whenever a target was sighted, the operator was directed to "take a picture" using the trigger on the joystick. This action recorded the viewing parameters of the current location to a log file. After finding a target, participants were instructed to further inspect target until they could accurately identify a discriminating feature marked on one side. Once this feature was found, participants were asked to take a second picture of the target object.

The environment (shown in Figure 23) loosely resembled a warehouse structure, with two levels connected by a ramp. The warehouse was comprised of a series of rooms that were arranged such that there was no obvious or continuous path that would cover the entire space. The closed layout meant that targets were generally not visible from a distance; navigation to each room was necessary to verify its contents. Aside from the target objects, the environment was void of non-architectural features.

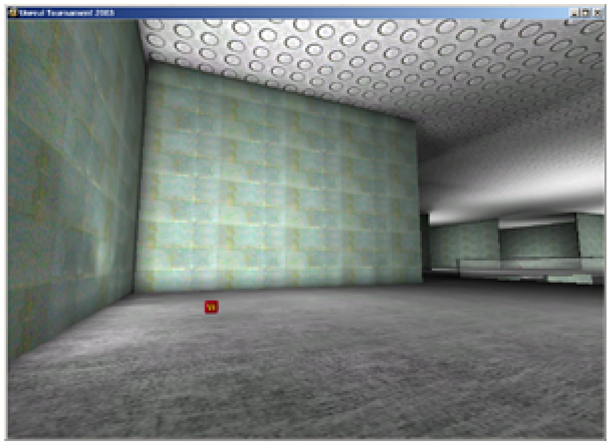


**Figure 23: Warehouse-like Environment**

Two conditions were compared in this experiment:

**Unassisted:** The operator was capable of moving the viewpoint forward-backward, pivoting, as well as panning and tilting the camera.

**Attentive:** In addition to the capabilities of the unassisted condition, attentive navigation was used to automatically adjust the viewing orientation. Whenever the viewer was in line-of-sight proximity to a target, the system would automatically pan the camera to focus on the object. This did not indicate which side of the target contains the identifying letter, only that a cube was present. Conversely, if there were no nearby target, the orientation of the camera would be automatically aligned with the direction of travel. The operator could override the recommendations by manually adjusting the pan/tilt, but whenever the robot was moving, control of the camera orientation was ceded to the robot.

*Apparatus*. The viewpoint was controlled using an enhanced function joystick. The main stick control was used to direct the position (forward and backward motion incrementally influenced the velocity, while side-to-side motion caused the view to pivot. The orientation of the camera was controlled using the hat-switch on the top of the joystick (Yaw was controlled by lateral movement, Pitch was adjusted by moving the hat switch forward and backward). The display was presented on a 21” monitor using 800x600 resolution.

*Procedures.* Prior to starting the task, participants were given verbal instructions on the objectives, and a demonstration of the controls. All subjects were required to confirm an understanding of the task and the controls by maneuvering through a separate training environment for several minutes and identifying at least one target object.

For the principle trial, twelve targets were evenly distributed throughout the environment. Targets consisted of a red cube marked on one side with a yellow, capital letter. Participants were advised that despite lighting conditions and cube placements, a discriminating letter was visible for every target in the environment (i.e. the letter was never face-down, or directly against a wall). Moreover, they were informed that the letters were unique and non-sequential. They were not told the number of targets in advance, although they did know that not all letters were represented.

Data were recorded in the form of a written list of all targets identified, as well as in an automatically recorded log file that tracked the position, velocity, heading and orientation. Entries were written to the log file nineteen times per second, allowing for a complete reconstruction of each session.

**Results**

One of the hypothesized benefits of attentive navigation is that nearby targets can quickly be pulled into view. This should alert the operator that a target is present and reduce the chances that it is overlooked. To test this hypothesis, the log files were parsed to find the amount of time that the operators spent near the various target objects. For the purposes of this study, “near” is defined by parameters used to trigger the camera reorientation. Near targets should have been visible to the operator provided that the viewing parameters were set correctly.

In the unassisted condition operators overlooked 15% of nearby objects, in contrast to the attentive condition where only 5% of nearby objects were overlooked, $t(24) = 2.51$, $p < .05$. Further analysis was done to assess how much time was spent near unidentified targets. As shown in Figure 24, the assisted technique spent an average of 24 seconds near unidentified targets, while the unassisted technique averaged 68 seconds. Thus the difference in the overall operator sensitivity is magnified by the relatively short exposure time for errors when using the assisted technique. Omissions likely occurred when quickly passing by the target, en route to some other destination. Conversely, the need for some guidance is underscored by the fact that operators fail to observe targets even when linger around the object for over a minute, in some cases as long as three minutes.

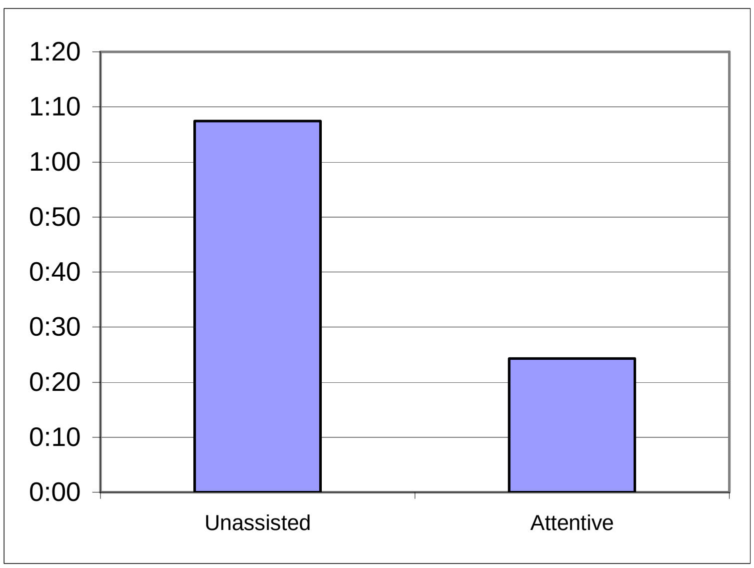



**Figure 24: Time Near Unidentified Targets (Experiment 3)**

Unfortunately, the added sensitivity afforded by attentive navigation did not translate into a better overall search performance in this case. There was no significant difference in the total number of targets found between the unassisted (6.77) and attentive (6.92) conditions. Given that the assisted techniques seemed to be superior at pulling targets into the field-of-view, why did the attentive condition not perform better?  The answer to this question can be found by studying the environmental coverage that each technique achieved.

From the movement logs, the average distance traveled for each participant was computed by totaling the Euclidean distance that the robot traveled over the course of the trial.  No major differences were found between or within groups.  This metric, however, does not distinguish between a robot that efficiently searched the environment and one that spent the duration of the experiment driving in circles.  To assess this distinction, a grid variance analysis was performed.  The environment was first partitioned into two 20x20 matrices (one for each level of the environment).  The positions recorded in the movement log were then allocated to the appropriate grid cells, providing a timed distribution of location.  Finally, the variance between cells was calculated to understand if the subjects uniformly searched the environment or loitered in specific regions.  The results of this process are presented in Figure 25.  The attentive condition show significantly higher grid variance $t(24) = 3.86$, $p < .01$, indicating that users of this technique did not explore as much of the environment as the unassisted viewers.

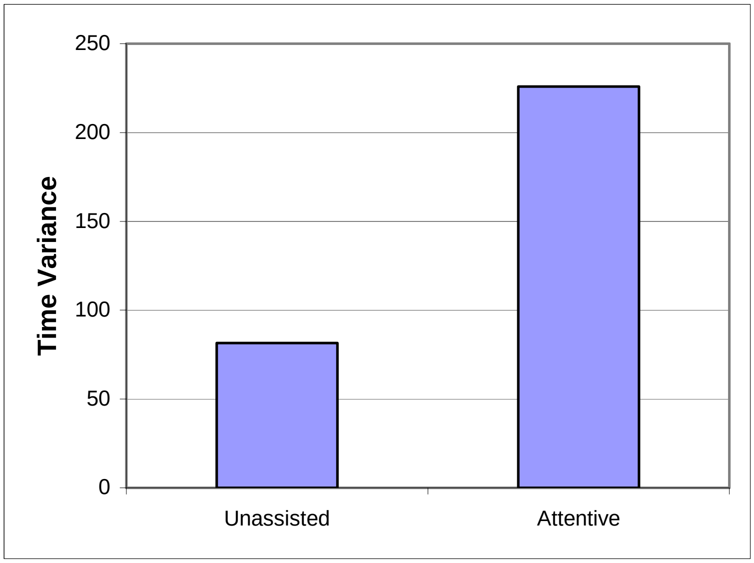



**Figure 25: Grid Variance (Experiment3**

The diminished coverage accounts for the absence of an expected performance boost. Essentially, the attentive condition performed admirably in the areas that it reached, but lacked comprehensive exposure of the environment.

### 4.2.4. Discussion and Recommendations

These experiments demonstrate an approach to exploring virtual environments that lies between the extremes direct presentation and unassisted navigation, capitalizing on the strengths of each. The results confirm H1; by sharing control of the viewpoint, the attentive techniques reduce the amount of interaction required to successfully extract the same amount of information. Therefore, the viewer is allowed to direct more attention on what is being observed during an exploration and less on controlling of the technique. At the same time, the viewer can still engage the display according to their needs. Unlike fully guided tours, the viewer makes decisions about the sequencing and observation time. Thus, the viewer retains a sense of self-determination, while the designer can ensure that certain information will be displayed.

At a very basic level, attentive navigation, using automatic gaze redirection, increased the probability that key viewpoints will be utilized, lending support to the acceptance of H2. This is witnessed by the inordinate number of missed objects and omissions registered by the unassisted navigators throughout the experiments. Naïve explorers in the first experiment may not have been aware that some objects would be elevated or contained inside other objects. Unassisted navigators remained oblivious, while attentive navigation successfully drew attention to these “hidden” elements. These benefits were evident with relatively small environments and should be expected to scale to large and more complex environments with even more pronounced effects.

The results also sustain H3. Attentive navigation helped viewers better understand the configuration of key elements, as well as their presence. These results from experiment 2 do not show an overall improvement in survey knowledge, but rather directed acquisition. This is a subtle, yet important distinction. Development of survey knowledge is largely dependent on repeated exposure to the landmarks in the environment. Unfortunately, given the complexity of realistic displays, an explorer may focus on irrelevant or redundant landmarks, prolonging the development of useful survey knowledge. The attentive navigation technique is effective at focusing the user's attention on significant elements in the environment, maximizing their exposure, and thus knowledge of their configuration.

Thorndyke and Hayes-Roth argued that motion greatly facilitates obtaining orientation-free survey knowledge [122]. One of the explanations they offered is that the navigator is being exposed to multiple perspectives. Attentive navigation takes this exposure a step further. By fixating the gaze on an object while moving past it, the viewer not only sees multiple facets of the object, but an extended context as well. In effect, the object remains stationary while the environmental background rotates around the object, maximizing the number of perspectives for the viewer. Since gaze fixation is built into the CMF, the viewer does not need to take any overt actions to take advantage of this optimized information.

The attentive techniques also foster the establishment of connections between elements of the scene, another key factor to learning survey knowledge. When using gaze-directed steering, viewers establish the intent to move by looking at an object they wish to explore. The same is true for this implementation of the attentive techniques. However, unlike unassisted navigation, as the viewer moves toward their objective, the viewpoint may shift to focus on another object. This redirection may forge a connection between the intended destination and the attended object: "As I walked toward the far wall, I looked to my left and saw X". Connections like this offer insight to the overall configuration of the environment.

H4 projected that attentive navigation would allow search tasks to be more thoroughly executed. In terms of the number of targets identified, this hypothesis was not validated by this set of experiments. However, the overall search effectiveness is dependent on the ability to maneuver the viewpoint efficiently and the sensitivity to nearby targets. Experiment 3 showed that attentive navigation apparently did improve the viewer’s sensitivity, but the maneuverability was seriously hampered. Anecdotally, in the first two experiments, some users of attentive navigation seemed to struggle for control of the gaze, frequently stopping to manually realign their gaze with the direction of motion. In the third experiment, this difficulty was manifest in the lack of coverage and quantified using the grid variance analysis. These results highlight the weaknesses of a strictly conditional control paradigm. Conflicts arose because the same camera view was used to highlight targets and make navigation decisions. In each of these three experiments, the assistance was toggled by motion in a specified region of interest. While the viewer was able to manually adjust the viewpoint, the feedback loop was not closed; the system did not learn from the viewers’ reactions. Thus, the system would blindly repeat its recommendation once the viewer initiated movement again. Clearly more advanced mechanisms and protocols for communicating the current knowledge state and information requirements need to be integrated into the system.

The fact that the benefits of the orientation assistance in the third experiment were completely negated by the moving difficulties is a point worthy of discussion. This is notably inconsistent with first two experiments where the mobility concerns were only manifest as complaints. The difference might extend from environmental disparities. In experiment 3, all of the target objects were reasonably visible and there was a lack of confounding non-targets. Experiment 1, by contrast, demonstrated an advantage in locating low-contrast and partially obscured targets. It stands to reason that the balance between sensitivity and mobility may shift depending on the complexity of the environment and the inherent conspicuity of the search targets. Further study might be able to determine if there is a threshold where the benefits of increased sensitivity outweigh the maneuvering difficulties.

### 4.3. Coordinated Assistance for Viewpoint Interaction

Drury et al argue that awareness between humans and automation is essential to productive collaboration [46]. This is witnessed in the previous three experiments where the only major difficulties were attributed to a poor communication between the recommendation system and the viewer. Given the open-loop feedback structure, the recommendation engine was not able to track the viewers' current goals and essentially acted independently of the viewer's changing information needs. It is possible that the interaction hardships from the conditional attentive navigation can eased by increasing the level of coordination. At this point, the objective is not to construct a system that that is capable of collaborating with the viewer as a peer. Rather the goal is to identify and implement key responses that can "intelligently" deal with complex information behaviors.

### 4.3.1. Coordination Mechanisms

Wrestling for control of the camera frequently took place at the boundary of these tasks: e.g. after the target had been identified and the viewer was ready to move onto the find the next target. At this point, camera fixation was no longer necessary and numerous participants indicated after the test that they would have done better if they could acknowledge that the target object had been identified, and remove it from future recommendation. While this solution might be ideal, unfortunately, this kind of feedback requires the system to maintain extensive object and environmental knowledge. This might be viable in constructed information visualizations, but to implement this solution in a sampled environment would require extensive real-time image processing and mapping capabilities that, according to Murphy, are currently lacking in most systems [93]. Therefore several alternative extensions to attentive navigation are outlined below.

**Manual Override** Many automation systems provide a manual override feature that allows the operator's judgment to supercede the decisions made by the automation. Section 3.3.3 warns of some of the difficulties that may arise when the viewer is solely responsible for modal operation of assistance. Moreover Yanco et. al. describe a series of errors that can occur when the operator mistakenly overrides the automation [141]. Rather than changing modes of operation, it may be beneficial for the operators to temporarily suppress viewing recommendations to allow them to execute a limited set of operations without interference.

**Contextual Override** It may be the case that the appropriateness of a recommendation can be inferred from the viewer's actions in the environment. Two very simple heuristics might inform the decision: Speed and Heading. If the viewer is traveling very quickly through the environment, he may not want to be distracted unless by a very important feature. In addition to the viewing orientation parameters, the CMF may contain a scalar that represents the strength of the attraction, proportional to the importance of the object. If the motion speed is less than the threshold, then the viewer's gaze could be redirected. If the motion speed exceeds the threshold, then the object may be ignored. Likewise heading may provide some insight to the viewer's intent; if the viewer is unequivocally moving away from the recommended viewing direction, it might be in the best interest to withhold the suggestion.

**Fixed Time** A viewer's interest in a particular feature may be temporal. Providing recommendations that expire after a short, fixed time may prove advantageous. The CubicalPath system, for example, attracts attention to features, allowing the viewer to understand the impact on navigation, and then releases the view back to the direction of motion, or other features of interest [9]. Toleration for interruptions may be garnered by ensuring that the interference will last only for a fixed duration. However, there are necessarily two parts to the fixed time strategy. If the system could identify the target that had expired, it could just as easily be explicitly acknowledged. Therefore, the second concern is how long the system should remain silent. It may be that the recommender system should reengage after a fixed amount of time. Alternatively, reactivation might be a function of distance traveled, reasoning that the viewer has moved sufficiently far away from the original target that new targets might be available.

**Parallel Viewing -** Hughes [70] proposed that the navigational subtasks could be split across two separate views. A fixed, forward-facing view effectively implements gaze-directed steering and allows the operator to understand the direction of motion (heading). The second screen allows the viewer to manipulate pan and tilt controls to alter the orientation of a different, co-located view. This perspective affords a rapid inspection of the environment, allowing the gaze direction to be independent of heading. If these views are simultaneously presented, the operator needs only to choose to attend to the screen that contains the information relevant to his current subtask.

Moreover, orientation recommendations, presented only on the second screen, would not interfere with the operator's ability to navigate.

Hypothetically, each of these techniques provides a higher level of coordination between the viewer and the automation for sharing control of the viewpoint than independent, conditional triggers. Since these methods rely on heuristics rather than explicit shared knowledge of the environment, it is not clear how effective they will be. Further evaluation is needed to gauge their relative impact.

### 4.3.2. Coordinated Control Evaluation

Each of these coordination heuristics was evaluated against the baseline, conditional gaze redirection paradigm following the same experimental framework and environment used in experiment 3 from the previous section.

The additional coordination between the viewers and the recommendation system should reduce the overall intrusiveness of the gaze redirection. This will free the viewer to complete a more thorough search of the environment. This coupled with the enhanced sensitivity afforded by attentive navigation should result in a better overall search performance, leading to the following extensions of hypothesis H4:

H4a: *Users of the Manual Override method will identify a higher mean number of targets in a search task than the conditional gaze redirection.*

H4b: *Users of the Contextual Override method will identify a higher mean number of targets in a search task than the conditional gaze redirection.*

H4c: *Users of the Fixed Time method will identify a higher mean number of targets in a search task than the conditional gaze redirection.*

H4d: *Users of the Parallel Viewing method will identify a higher mean number of targets in a search task than the conditional gaze redirection.*

**Method**

*Participants*. 57 paid participants were recruited from the University of Pittsburgh undergraduate community. 14 subjects were randomly allocated to one of four conditions with one additional participant used to balance data from a previous condition. Previous studies have resulted in effect sizes ranging from 15% to 20% for mean target identification with standard deviations approaching 25%. Anticipating similar values for this study, for the statistical tests to achieve the desired power value of 0.8, a minimum of 14 subjects per condition is required. Given a dependence on vision and identification of color objects, it was stipulated that participation required a self-report of normal or corrected-to-normal color vision.

*Stimuli.* The same protocol, environment and apparatus will be used from experiment 3 described in Section 4.2.3 to evaluate four additional conditions:

**Manual Override:** Recommendations were made on the same basis as previously expressed – centering the view on targets that are within line-of-sight proximity to the viewer. To override the recommendations, viewers were required to hold down a button located at the base of the joystick. This had the effect of ignoring any recommendations as long as the button is depressed; all camera-panning motions were the responsibility of the viewer.

**Contextual Override:** Two contextual parameters were suggested for overriding the recommendations: speed and relative heading. Since none of the targets have priority in this experiment, a constant threshold could be used for all of the target objects. The effects of speed and relative heading were combined as shown in Figure 26. If the target is within 45°of the heading, the recommendation was displayed regardless of speed. If the target is between 45° and 90° of the heading, the recommendation was shown if the speed is less than 75% of the maximum speed. Likewise, the speed must have been less than half of the maximum to make a recommendation for targets between 90° and 135° if the heading. Targets that were greater than 135° from the heading, were not recommended.

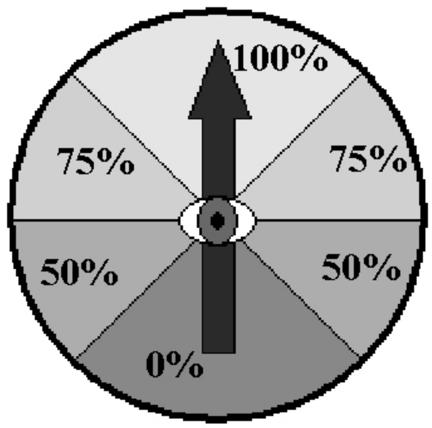



**Figure 26: Range of Recommendations**

**Fixed Time:** The selection of the time interval for the fixed time is critical. If the duration were too short, the recommendation may help only with alerting the viewer to the presence of the target, but not with the fixation. If the duration were too long, then the viewer would start to compete with the automation for control of the camera. An evaluation of the movement logs for the conditional gaze redirection from experiment 3 indicates that it took an average of 20 seconds per target before identification could be made. Therefore, after 20 seconds of fixating on a target object, the recommendations expired. For this experiment, 12 seconds was selected arbitrarily as the reactivation time.

**Parallel Viewing:** This treatment provided two separate views, each displayed on a distinct monitor. One monitor represented a fixed camera – effectively implementing gaze-directed steering. The second monitor displayed the recommendations of the attentive navigation system. The operator had the ability to pan and tilt the second camera, temporarily overriding the recommendations, but automation resumed whenever the viewpoint is moved.

Data previously collected from the conditional attentive navigation is included as a baseline for the analysis. To ensure equal sample sizes for the analysis, one additional participant was added.

*Procedures.* The experiment was conducted in accordance with the procedures laid out for experiment 3. Prior to starting the task, participants were given verbal instructions on the objectives, and a demonstration of the controls. Details of the appropriate coordination strategy, as well as a physical demonstration were also given to the participants. All subjects were required to confirm an understanding of the task and the controls by maneuvering through a separate training environment for several minutes and identifying at least one target object.

The same training and test environments from experiment 3 were used to execute this evaluation. For the principle trial, twelve targets were evenly distributed throughout the environment. Targets consisted of a red cube marked on one side with a yellow, capital letter. Participants were advised that despite lighting conditions and cube placements, a discriminating letter was visible for every target in the environment (i.e. the letter was never face-down, or directly against a wall). Moreover, they were informed that the letters were unique and non-sequential. They were not told the number of targets in advance, although they did know that not all letters were represented.

Data were recorded in the form of a written list of all targets identified, as well as in an automatically recorded log file that tracked the position, velocity, heading and orientation. Entries were written to the log file nineteen times per second, allowing for a complete reconstruction of each session.

**Results**

The average number of targets identified for each condition is presented in Figure 27. A single-factor analysis of variance (ANOVA) was performed on the overall number of targets identified, and a main effect was found for the camera control technique used $F(4,65) = 2.59$, $p<.05$. Using the

Bonferroni post-hoc procedure to compare the coordinated conditions to the conditional attentive navigation, it was further verified that the users of the manual override and parallel viewing conditions consistently identified more targets: $t_{prot}(65) = 2.43$, $p<.05$ and $t_{prot}(65) = 2.92$, $p<.01$ respectively. The contextual override and timed override did not perform reliably better than the conditional attentive navigation

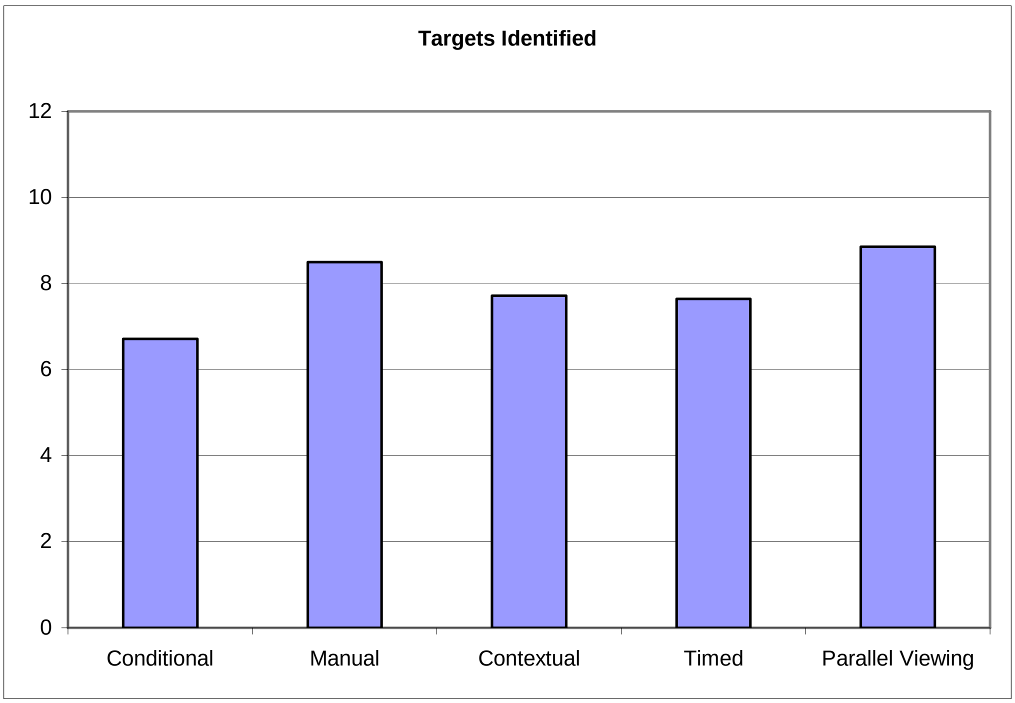



**Figure 27: Time Near Unidentified Targets (Experiment 4)**

There are two major factors that affect the overall effectiveness of a techniques search capabilities: sensitivity and mobility. Sensitivity attempts to quantify the ability of the viewer to discriminate useful views from noise; if they are in a position to receive the information do they actually acquire it. Mobility on the other hand measures the viewers' ability to achieve broad coverage of the environment, increasing the chances that they will encounter useful information in their search.

The sensitivity measures the likelihood that the viewer would identify nearby targets. An average of 6% of the nearby targets were overlooked across all of the attentive conditions; there was no reliable distinction across the different techniques $F(4,65) = 0.40$. By comparison, recall that the unassisted condition missed nearly 15% of nearby targets. This suggests that providing assistance, regardless of the coordination strategy is likely to increase the sensitivity to nearby elements of interest.

To determine the impact of the interaction condition on mobility, an analysis of grid variance was performed to compare the environmental coverage of the various control techniques. As in experiment 3, the environment was partitioned into two 20x20 matrices, one for each level of the environment. The positions recorded in the movement log were then allocated to the appropriate grid cells, providing a timed distribution of the viewing location. The variance between the cells was calculated to understand if the subjects uniformly searched the environment or loitered in certain regions. The results of this process are found in Figure 28. The conditional attentive navigation had a significantly higher grid variance than all the coordinated techniques (t-tests, $p<.01$). Moreover, the timed override also had significantly higher grid variance than the other coordination techniques: vs. manual override $t(26) = 2.29$, $p < .05$ ; vs. contextual override $t(26) = 1.74$, $p < .05$; vs. parallel viewing $t(26) = 1.74$, $p < .05$.

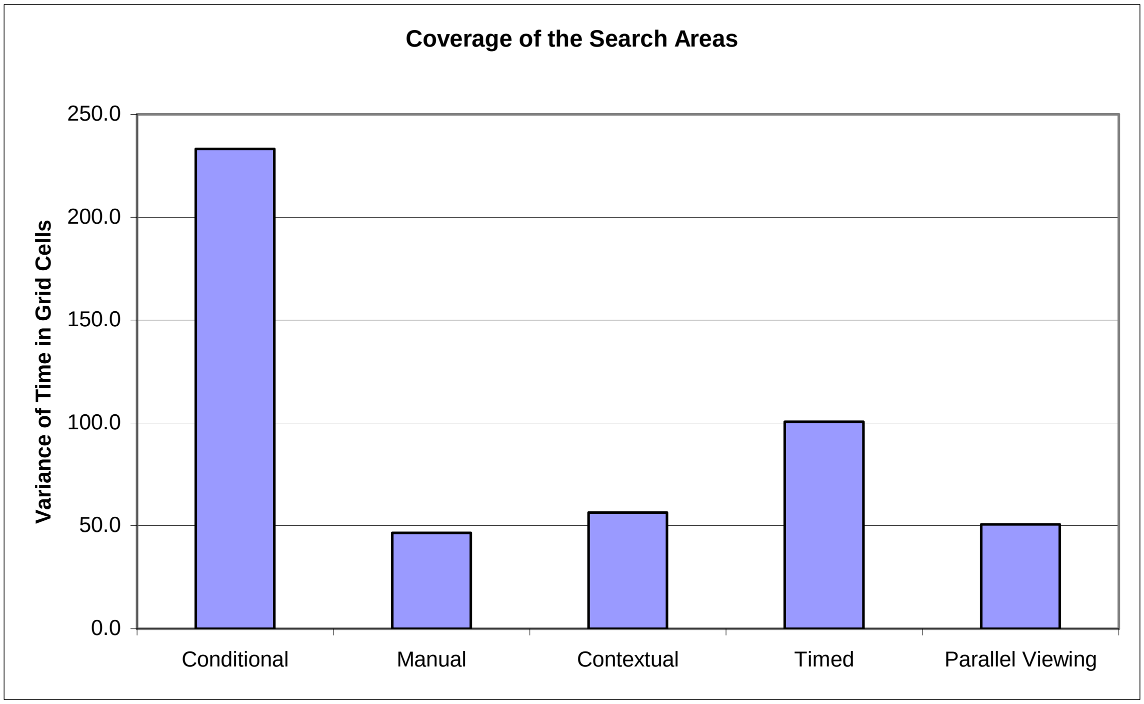



**Figure 28: Grid Variance (Experiment 4)**

Grid variance gives a broad overview to the viewpoint mobility. To better understand the movement patterns of the different conditions, further analysis was done with respect to actions taken when nearby targets. Recall that the viewer was directed to take two distinct pictures of every target encountered. The timing of these pictures outlines three distinct types of interaction with each target. The first period is from the time that the subjects are near the object until they indicate that they have seen it. Interaction at this phase is characterized by attempts to find the target. . The time span between the two pictures is captures by the subject's efforts to identify the object. In the final stage, the subject is still near the target object, but they have no reason to remain, and should be moving away. The breakdown of interaction techniques over these three time intervals is presented in Figure 29.

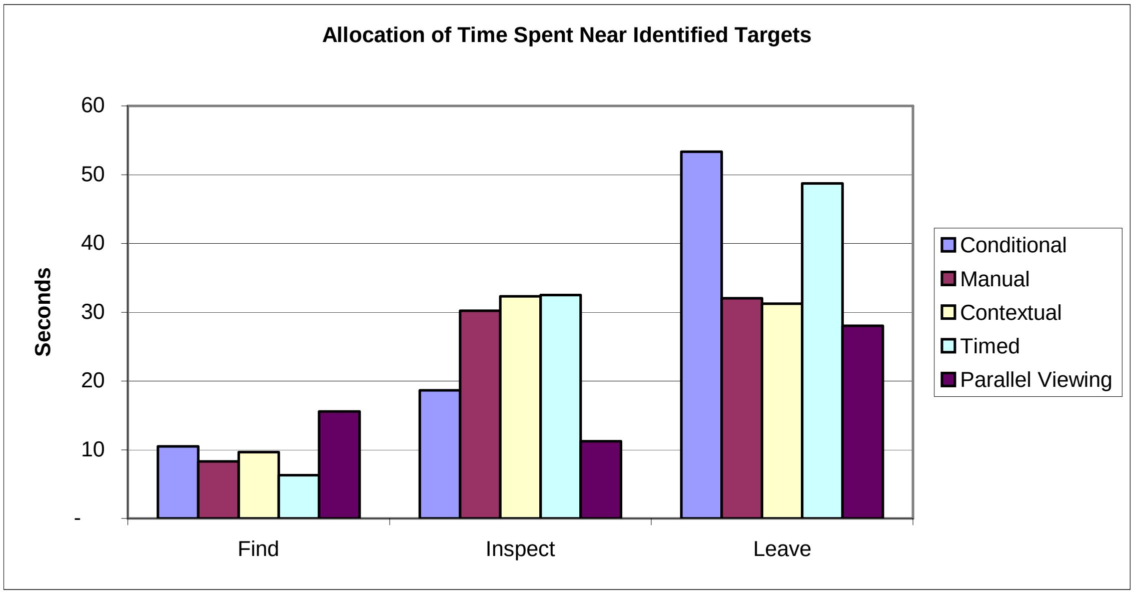



**Figure 29: Time Allocation Near Identified Targets (Experiment 4)**

For all three subtasks, there was a high degree of variability introduced by the disparities in individual target difficulty. For example, the 'B' target was at the end of a dead-end corridor, with the letter facing the open path. Whenever the viewers found this target, they could almost immediately identify it as well. Conversely, it took the viewers an average of nearly a minute to identify the 'Z' target because it was poorly lit and its awkward position relative to a nearby wall. Because of this between-target variability, a pair-wise analysis is used to determine the find, inspect and leave times for each target across the different control conditions.

The parallel viewing condition was significantly slower in the find segment than conditional attentive navigation for 10 out of 12 targets with a mean difference of 5 seconds (33%), $t_{pair\text{-}wise}(11)=2.63$, $p < .05$. There were no reliable differences in the search time for the other conditions. The extended time for the parallel viewing condition is likely caused by the fact that the viewers have their attention split across two screens; they may only be periodically monitoring the screen with the recommendation, causing a slight delay in response time.

For the inspection phase, the parallel viewing condition performed significantly better than conditional attentive navigation, identifying 10 out of 12 targets faster: $t_{pair\text{-}wise}(11) = 2.21$ $p < .05$. In this case, the split-screen approach seems to have benefited the viewer; they could instantly toggle between navigation and inspection without having to take any explicit action. It may appear from Figure 29 that the manual, contextual and timed override had longer inspection times than conditional attentive navigation, but there were no reliable differences for these techniques.

Efforts to leave a target after it had been identified were substantially hampered for conditional attentive navigation compared to the manual override (9 of 12 targets, $t_{pair\text{-}wise}(11) = 3.06$ $p < .01$ ); contextual override (10 of 12 targets $t_{pair\text{-}wise}(11) = 4.52$ $p < .01$ ); and parallel viewing (11 of 12 targets $t_{pair\text{-}wise}(11) = 4.25$ $p < .01$ ). Likewise the timed override condition also had a relatively difficult time moving away after the target was identified (manual override: $t_{pair\text{-}wise}(11) = 1.80$ $p < .05$, contextual override: $t_{pair\text{-}wise}(11) = 2.69$ $p < .05$, parallel viewing: $t_{pair\text{-}wise}(11) = 2.33$ $p < .05$).

Working from the premise that the conditional technique's mobility (and hence search effectiveness) was stunted by its intrusiveness, it is instructive to understand the extent to which the viewer engaged in overriding the system's aid. A breakdown of the relative times that the recommendations were active reveals two interesting trends. As seen in Figure 30, reasonably distinct bands suggest that the relative duration of the recommendations depended on technique being used. In fact, this distinction stands up to statistical scrutiny. The average time spent overriding recommendations varied by interaction technique: Manual override = 50%, Contextual Override = 38%, and Timed Override = 28%, $F(2,33) = 13.31$, $p<.001$. The parallel viewing approach was omitted from this analysis. Technically, the viewpoint recommendations are always active and the override occurs when the viewer looks at the fixed-orientation screen. Since this experiment did not track which screen had the viewers' attention, there is no way to compute a comparable metric. However, it may be reasonable to assume that the override profile would be similar to the manual condition since both conditions were based on viewer-initiated overrides.

The second interesting insight from Figure 30 can be derived from the shapes of the concentric rings. The shape of the timed override curve is essentially circular, indicating that that the duration of the recommendation was uniform, with very little variance based on the intended target. While the timed override reduces the fixation on targets relative to the conditional attentive technique, it does not factor in the viewers' information needs in this process. By contrast, the shapes of the manual and contextual override conditions are more jagged, capturing the variation required for targets that were easier or harder to identify. It is encouraging that the contour of these two curves is similar. This suggests that the heuristics used to deduce the needs of the viewer in the contextual condition were a reasonable match to explicit actions taken by the viewer in the manual condition.

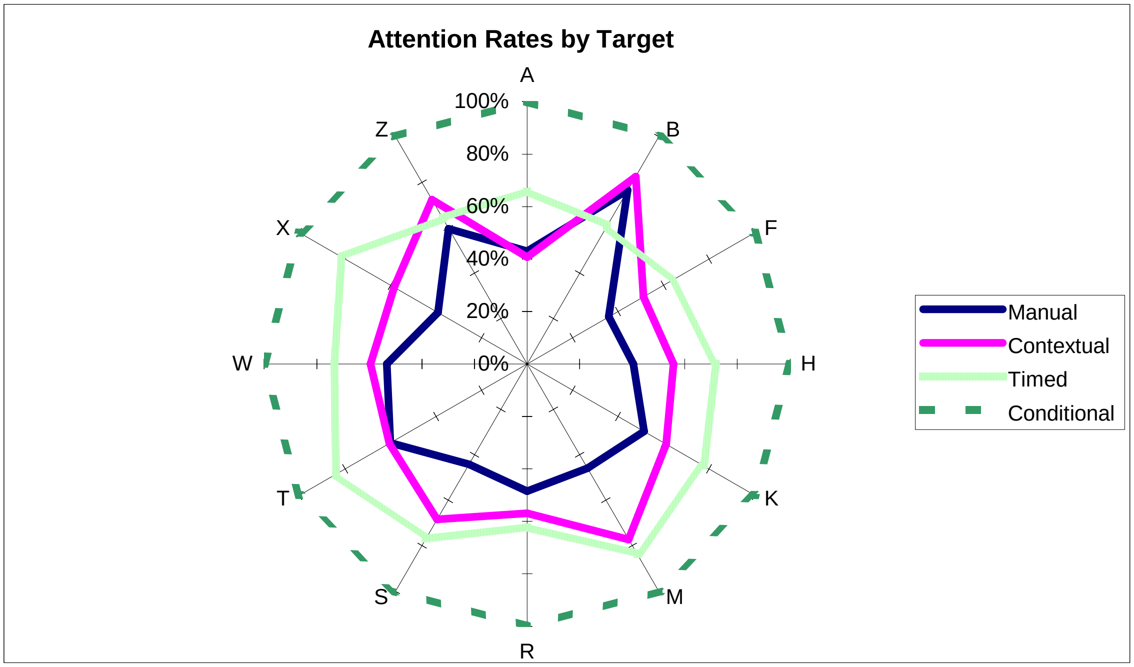



**Figure 30: Relative Assistance Provided by Techniques**

The results from this experiment are summarized in Table 6.

**Table 6: Summary of Results from Experiment 4**

| | Search Effectiveness | Sensitivity | Mobility | Relative Intrusiveness |
|---|---|---|---|---|
| Manual Override | Increased | High | High | 50% |
| Contextual Override | No Change | High | High | 62% |
| Timed Override | No Change | High | Low | 72% |
| Parallel Viewing | Increased | High | High | Unknown, presumed Low |

### 4.3.3. Discussion

The results presented above support hypotheses H4a and H4d – indicating that manual override and parallel viewing are two coordination strategies that can reduce the control struggles that cause problems with attentive navigation. The data do not support H4b (contextual override) and H4c (timed override). This does not necessarily mean that these two techniques are not viable means of coordination in assisted viewing, only that our study failed to demonstrate their potential value. The discussion below is broken down to analyze each of the techniques evaluated in this study individually. The strengths and weaknesses brought out by this study are reviewed along with additional speculation on further refinements that can be made.

#### Manual Override

The addition of a manual override option enhanced the performance of conditional attentive navigation. Viewers using this condition were able to identify more targets and explore a wider area of the viewing environment. These two results likely stem from the fact that the operator was able to explicitly communicate that they had finished with a target and override the recommendations to quickly leave and move onto another.

Since the viewer determines when to receive recommendations, and when to override, there was initial concern that overconfident viewers would use the override excessively, causing them to miss targets. While there was no difference in the overall sensitivity, there is some evidence to support this concern. The manual override condition spent an average of 35 seconds longer near targets that were not identified than the other coordinated conditions $t(55) = 2.51$, $p < .01$. In fact, the time spent near unidentified targets was not statistically differentiable from the performance of the unassisted technique $t(25) = 0.78$. These targets were not found largely because of overridden viewing recommendations. For targets that were not identified, the viewing recommendations were overridden 66% of the time (contrasted with a 50% override rate for identified targets). It is important to keep in mind that while the recommendations for these targets were active 34% of the time, it was not guaranteed that the target was actually displayed on the screen – the operator could have engaged the override while the automation was still adjusting the camera.

The design principle that the override was maintained by the persistently explicit act of holding down a button helped to manage the frequency and duration of the overrides. Each override action lasted an average of less than 8 seconds. If the viewers were able to disable the recommendation engine entirely, it is more likely that they would forget to reactivate it – leading to the modal problems described in sections 3.3.3 and 4.3.1

**Contextual Override**

This experiment was unable to show the value of the contextual override Interestingly, this condition was aligned with the manual override and the parallel viewing in terms of coverage and actions taken nearby target objects, suggesting the that the gaze redirection did not have a negative impact on mobility. Moreover, the contextual override was highly sensitive to nearby targets, missing less than 5%. Unfortunately, these benefits did not translate into the ability to reliably identify more targets.

Lacking clear impediments to this technique, it is entirely possible that the experiment simply was not strong enough to distinguish this interface from the conditional attentive camera. A closer look at the data for these conditions shows that the mode for contextual override was 9 targets identified (5 out of 14 participants), while the mode of the attentive navigation condition was only 6 targets (5 out of 14 participants). In fact, the contextual override target identification results were negatively skewed (-.72) and the attentive condition results were positively skewed (.92). While both of these values fall within the acceptable range of 2 standard errors of skewness for this sample size (1.31), it might have tipped the balance toward a type II error. Given the potentially skewed sample of this dataset suggests that the sample size may have too small to show a difference or there is a potential ceiling effect that is bounding the upper end of the distribution. It is possible that re-running the experiment with more participants would provide a more accurate sample. Likewise, expanding the number of potential targets could allow for a larger effect size and thus reveal an undetected difference.

Despite the lack of quantifiable evidence for the contextual override, the heuristics used for overriding the recommendations seemed quite effective. As shown in Figure 30, the difference between the relative times for manual override and contextual override seem to be a nearly linear offset. It is possible that fine-tuning the parameters for the contextual override may allow the automation's to more accurately deduce the needs of the viewer and achieve the desired performance boost.

Finally, there was an interesting and unintended side effect from the contextual override interaction. As it was conceived, the viewers were to partially communicate their intentions by modulating their speed. Moving slowing would indicate that they were receptive to viewpoint redirection, while moving quickly would signify other intentions. For this environment and objectives, viewers spent most of their time moving at near top speed. This means that in practice, the motion was not graduated, but rather moving or not. Thus, Figure 31 provides a more accurate characterization of the interaction that is taking place. When moving, the recommendations are active within 45° on either side of the motion vector (region A). When the viewer stops moving, if there is a target between 45° and 135° (region B), the orientation of the camera will be automatically adjusted. Then when the viewer starts moving again, recommendation would be overridden again and the gaze would shift back to the motion vector. If the target were located in region C, the camera orientation is unaffected. We observe that using the contextual override in this manner approaches Wernert and Hanson's conditional policy of non-interference where automation would only occur when the viewer is not issuing any direct commands [137].

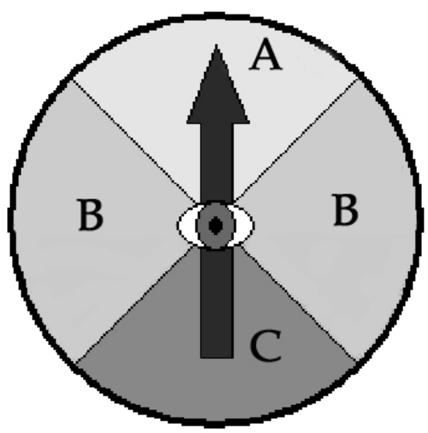



**Figure 31: Alternative Binary Contextual Override**

### Timed Override

In contrast the contextual override condition, there seemed to be a more fundamental flaw in the timed override condition. Rather than the optimistic statement that the experiment failed to prove this technique's worth, there is actually evidence that this interface was problematic. Like the conditional attentive navigation, there were significant problems leaving identified targets and overall search coverage that had a negative impact on the viewer's ability to complete the task.

Earlier it was noted that establishing the time parameter would be a critical factor in the success of this approach. Unfortunately, only 61% of the targets were identified in the before the time elapsed and the recommendations were turned off. Moreover, 12% of identified targets exceeded the time boundaries more than four times. This level of toggling back and forth can prove quite disruptive to the operation and likely contributed to the poor performance.

After completing this experiment, it is no longer believed that improving the performance is simply a matter of finding the appropriate timing parameter. As stated in the results section, there was a large variability in the time needed based on the difficulty of the target. Figure 32 shows the distribution for the average time taken to find and inspect each of the target objects. Note that almost all of the letters were identified while the recommendations were active (denoted in the figure by the shaded bands). Some of these targets could have been identified earlier had the recommendation not been overridden prematurely. Nonetheless, the average time to find and identify targets range from 5 seconds to 87 seconds. It is unlikely that a single, fixed duration could be found to effectively assist with all possible elements of interest. This does not mean that the timed override should be abandoned outright. If there were an environment that consisted of a known, homogeneous constellation of interest points, the timed override might yet have value – however, the plausibility of that scenario and the verification of the techniques effectiveness are left to be determined. Alternatively, it may be possible to store the duration of attention as an attribute in the CMF. Mission critical or difficult objects may be assigned a long fixation time, while secondary or easy to identify targets may be assigned a shorter duration.

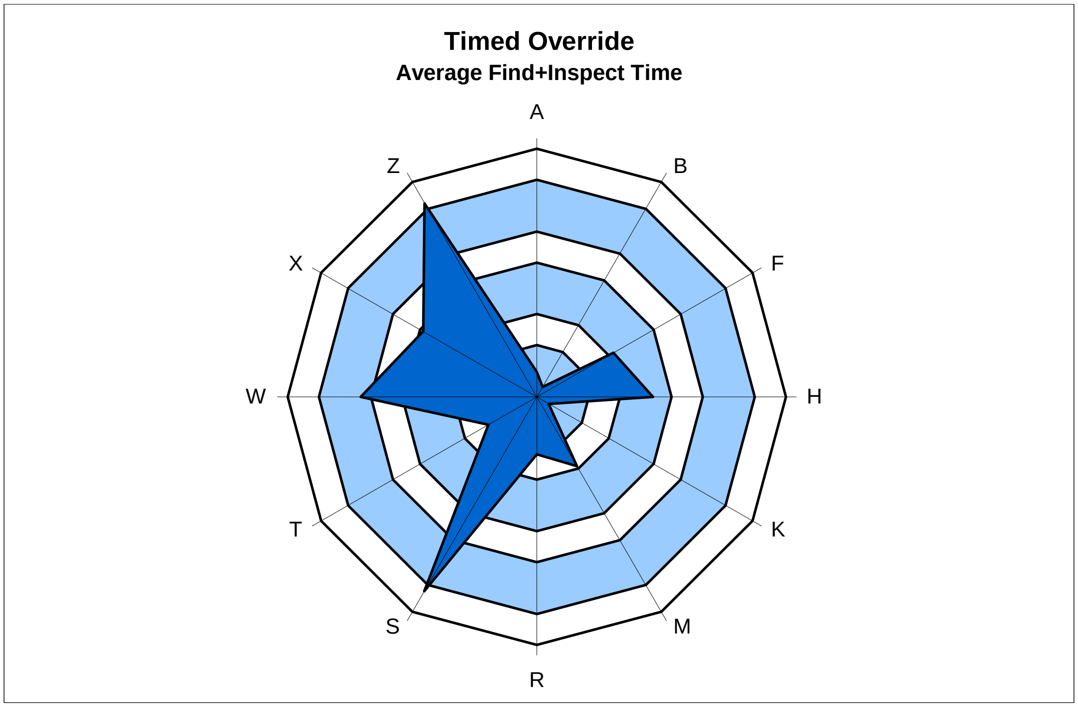



**Figure 32: Time Needed to Identify Letters for Timed Override**

## Parallel Viewing

This experiment showed that the parallel viewing condition offered better search performance, and easier mobility through the environment than conditional attentive navigation. It could be argued that the benefits of the parallel viewing can be attributed to the increase in information being transmitted. With two separate views into the dataset, this condition is sending back potentially twice the visual information than the single-view approaches.

To address this issue, a follow-on condition was explored to assess the impact of multiple viewpoints. Adhering to the same stimuli and protocol for the other conditions, 13 additional participants completed the search task using a modified version of the parallel viewing condition. In this new version, the viewer still interacted with two views. The first was the same fixed, forward-facing camera used in the previous condition. The second view allowed the viewer to independently control the orientation of the camera through pan and tilt operations. Unlike the attentive parallel viewing condition, the orientation of the camera was strictly controlled by the viewer; there was no automation to make the recommended adjustments to the camera orientation.

The unassisted parallel viewing condition found an average of 7.5 targets. This was not significantly different from the single-camera unassisted condition (6.8 targets, $t(24) = 0.80$). It was, however, significantly lower than the 8.8 objects found by the assisted parallel viewing, $t(25) = 1.71$, $p < .05$. This result suggests that neither the conditional attentive camera nor the use of multiple cameras alone improve general search effectiveness. However the conjunction of these two factors reliably produces better results.

# 5. Conclusions

The study of visualization spans a wide range of methodologies, theories and applications aimed toward one goal: augmenting the human capacity to process and digest information. The recent advent of interactive visualization introduces the potential to rapidly explore relationships, confirm hypotheses and comprehend large datasets like never before. To begin to harness this potential, this research has systematically explored the roles of humans and automation in interactive viewpoint manipulation in 3D representations. This section summarizes the findings of this research, offers perspective by commenting on its significance and speculates about future directions and unanswered questions.

## 5.1. Summary

Viewpoint interaction was specifically targeted as a critical operation to understanding a wide array of applications ranging from abstract configurations, to scientific models, to teleoperation in remote environments. An overview of the literature reveals that the physical and cognitive operations associated with viewpoint manipulation requires considerable effort. Moreover, control difficulties are exacerbated by the nature of the displays that often contain extensive noise and sub-optimal views. These problems motivate the need for viewing techniques that can assist the viewer to focus on relevant information, while minimizing efforts to adjust the viewpoint.

To better understand the complexities of assisted viewpoint control a framework of key design considerations was constructed. First, it is important to understand the nature of the assistance. Analyzing typical interactions suggests the need for guided positioning – moving the viewpoint to critical locations, and guided orientation – helping the viewer to look in the right direction, or both. Second, there is a discussion about how the assistance is generated. To provide automated assistance, the system needs to be able to extract knowledge from the environment in real-time, or have it embedded after a preprocessing analysis. Finally, the framework addresses the methods for communicating viewing recommendations to the viewer. These include annotation and external cues, restriction and prioritization.

An example system was implemented and evaluated specifically to study the effectiveness of interaction methods that actively provide viewing assistance. Attentive navigation redirects the viewing orientation to focus of critical elements of the display based on the location within the environment. Three user-studies had promising results, specifically revealing a successful reduction in controller complexity, and improvements to the transfer of landmark and survey knowledge. The impact of attentive navigation on search effectiveness was inconclusive.

Studying the results of the user-studies led to a second iteration in the design of attentive navigation. It was hypothesized that the conditions that triggered the recommendations were too broad, resulting in intrusive adjustments to the viewing orientation and a decline in viewers' overall mobility. This theory prompted the development of several mechanisms for overriding recommendations and providing more coordination between the viewer and the automation. Four such enhancements were implanted and evaluated. This user study resulted in two of the techniques (manual override and parallel viewing) significantly improving the viewers' search capabilities. Another approach (contextual override) did not reliably enhance the search effectiveness, but there was evidence to warrant further investigation. The final technique (timed override) proved insufficient at improving the viewers' search capability.

## 5.2. Contributions and Significance

This research makes both theoretical and concrete contributions to the interactive visualization community. The problem of viewpoint control is a critical issue that has an impact the entire spectrum of 3D visualizations; if you are not looking at the right information, the visualization does not serve its function. This is as true for a constructed cone-tree of hierarchical data as it is for a video feed from a teleoperated rover on the surface of Mars. The design considerations, while short of a comprehensive list, provide the basis for discourse as future systems are developed. Previous attempts have been ad-hoc and difficult to compare and leverage off each other. Working with the framework provided in section 3, exposes some of the major concerns for research in this field and can guide future development in a more systematic way.

These design considerations were used to craft attentive navigation as a viable technique for viewpoint assistance. Specifically, attentive navigation provided prioritized viewing orientations based on expert mark-up of the viewing environment. The findings of the user evaluations make it possible to put this technique into practice because it demonstrates several qualities that are highly desirable in interactive visualization:

- It allows sufficient interaction to completely explore a 3D environment with standard input devices without the complexity of specialized 6DOF controller or awkward control mappings. This could hasten the proliferation of 3D visualization applications to standard workstations.
- Viewers are not constrained by scripted interaction, they benefit from a sense of self-determination without the cognitive burden of maintaining excessive viewing parameters.
- Employing this technique results in demonstrably better landmark recognition, sensitivity to key features, directed survey knowledge and improved search effectiveness.

The experiments presented in this dissertation implemented and evaluated attentive navigation in the context of a *modeled* environment. As such, the findings are most directly relevant to scientific visualization research. However, the environments for experiments 3 and 4 were constructed using USARSim, which has been shown to provide interaction characteristics similar to teleoperation interfaces [129], suggesting that attentive navigation might also serve that class of visualization as well. It is reasonable to expect that information visualization applications may benefit from attentive navigation as well. Although there were no explicit tests in this study, similar techniques have been evaluated in the context of constructed displays [42]. However, given that the spatial layout of these displays is somewhat arbitrary, the viewer might be better served by restructuring the representation rather than extensive viewpoint manipulation – this remains an open question

While attentive navigation may be ready to plug-in to some scenarios, there are some concerns that considerably limit the potential applications. Specifically, the pre-authored nature of the CMF is incompatible with most teleoperation activities, where there is little or no foreknowledge of the environment. However, the findings of this research can be can also be interpreted broadly, dealing with assisted viewpoint interaction issues that are independent of a specific technique. Regardless of how the recommendations are generated, this study provides some valuable insight to how operators react to automation that is capable of asserting its own agenda. The differences between the conditional attentive navigation, and the coordinated approaches underscore the importance of awareness between the viewer and the automation. When the automation blindly made its recommendations, it proved too intrusive and negated the potential benefits of the assistance. While the override techniques provide more coordination with the automation than the initial conditional location, they still fall short of the level of discourse imagined by true mixed-initiative systems. For example, it has been suggested by other researchers that the automation should be infused with social mores to limit the automation to "appropriate" interruptions [51]. Implementing such a policy for viewpoint control would push the current bounds of artificial intelligence technology and would likely increase the complexity of the automation by an order of magnitude. By comparison, the relatively minor heuristics demonstrated in this study significantly improved coordination with very little computational expense. Rather than seeking automation that is outright "intelligent", designers would do well to imbue technology with small bits of intelligence that can make the largest impact.

### 5.3. Future Directions

Numerous research opportunities exist for increasing the state of knowledge with regard to assisted viewpoint interaction. This work has shown attentive navigation to be effective at certain aspects of assisted viewpoint interaction, however the evaluation covers only a small portion of the design considerations laid out in Section 3. Further work should be done to elaborate on the relative value and applications of other combinations from this framework. Given the nascence of this field, development of new techniques for assisting viewpoint manipulation is still needed. While new techniques would be welcome, attentive navigation provides a powerful engine for exploring these issues. For example, section 4.1 describes how attentive navigation can be used to provide guided positioning, however no formal evaluation has ever been conducted. Likewise, the vectors stored in the CMF can be used to power external cues and annotations.

An additional topic for future exploration addresses the changing needs of the viewer. It is one thing for the automation to make a binary decision to interrupt the viewer based on fear of intrusion. It is quite another to offer partial or occasional advice with an understanding that the viewer may benefit from the struggle of discovery. Naive instructional devices often rely on students learning from mistakes over numerous repetitions in order to foster understanding – a technique known as "drill-and-practice". An alternative approach, commonly referred to as "scaffolding" describes a collaborative process that allows a student to complete tasks that might ordinarily be too complex for them. This concept is built from the work of Vygotsky, who proposed that instructors should manage learners' experiences by giving them portions of the task that are within their 'Zone of Proximal Development' [128]. Over time the students take a more active role in the completion of the task until they are able to complete it independently. It has been argued that this kind of pedagogy has been successfully built into other online tutoring systems [66], and might be equally effective when applied to reasoning with visualizations The scaffolding process can be achieved by refining the amount of control that the viewer exerts, ranging from passive to complete. Initially, the viewer can watch as the system completely automats the viewpoint to point out the important features. As the viewers begin to understand what is significant, they can attempt to interact with the display with less intervention from the system. When the viewer has obtained a good understanding of the organization in the environment, they might only need the hints available by request. Ultimately, when the viewer has achieved the desired learning, they can freely navigate without the assistance of the system. Using this approach it may be possible to provide a more structured approach to learning by progressively granting more viewpoint control to the viewer.

Trust is another critical element to viewpoint recommendations that was not explored in study. “A decision aid, no matter how sophisticated or ‘intelligent’ it may be, may be rejected by a decision maker who does not trust it.”[91]. Trust is a fairly broad concept that is based on a number of factors including: consistency of behavior, demonstrated competence, and faithfulness to obligations [80]. If the system is mistrusted, not only will the potential benefits be lost, but decision makers may go out of their way to bypass the guidance [91]. This reinforces the idea that no advice is often better than bad advice. Further work needs to be done to assess whether or not the benefits recorded in these studies will hold up in the face of occasional bad advice.

## 5.4. Closing

Ultimately visualization will play an important role in dealing with the wealth of information facing people today. Assisted viewpoint interaction technology will ensure that meaningful information is not only displayed, but consumed as well.

## 6. References


[1] Anderson, J.R., *Cognitive Psychology and its Implications*. Fifth ed. 2000, New York: Worth Publishers.

[2] Arthur, E., P. Hancock, and S. Chrysler. *Spatial Orientation in Real and Virtual Worlds*. in *Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting*. 1993. Seattle, Washington.

[3] Arthur, E., P. Hancock, and S. Telke. *Navigation in Virtual Environments*. in *Proceedings of the SPIE – International Society for Optical Engineering*. 1996.

[4] Azuma, R., *A Survey of Augmented Reality*. Presence: Teleoperators and Virtual Environments, 1997. **6**(4): p. 355-385.

[5] Bajscy, R., J. Kosecka, and H. Christiansen, *Discrete Event Modeling of Navigation and Gaze Control.* International Journal of Computer Vision, Special Issue on Qualitative Vision, 1995. **14**(2): p. 179-191.

[6] Baker, M.P. and C.D. Wickens. *Human Factors in Virtual Environments for the Visual Analysis of Scientific Data* 1995: NCSA-TR032 and Institute of Aviation report ARL-95-8/PNL-95-2

[7] Bares, W., T. Somying, and S. McDermott. *A Model for Constraint-Based Camera Planning*. in *In Smart Graphics: Papers from the 2000 AAAI Symposium*. 2000: AAAI Press.

[8] Beaten, R., R. DeHoff, N. Weiman, and P. Hildebrandt. *An Evaluation of Input Devices for 3-D Computer Display Workstations*. in *Proc. of SPIE-The International Society for Optical Engineering*. 1987.

[9] Beckhaus, S. Dynamic Potential Fields for Guided Exploration in Virtual Environments. Doctoral Dissertation. Computer Science, University of Magdeburg. 2002

[10] Beckhaus, S., F. Ritter, and T. Strothotte, *Guided Exploration with Dynamic Potential Fields: The CubicalPath System.* Computer Graphics Forum, 2001. **20**(4): p. 201-210.

[11] Bederson, B.B. and J.D. Hollan. *Pad++: A Zooming Graphical Interface for Exploring Alternate Interface Physics*. in *Symposium on User Interface Software and Technology*. 1994. Marina Del Rey CA.

[12] Bell, B., S. Feiner, and T. Hollerer. *View Management for Virtual and Augmented Reality*. in *ACM Symposium on User Interface Software and Technology*. 2001. Orlando, FL.

[13] Bertin, J., *Graphics and Graphic Information Processing*. 1981: Walter de Gruyter.

[14] Bowman, D. Interaction Techniques for Common Tasks in Immersive Virtual Environments: Design, Evaluation and Application. Doctoral. Computer Science, Georgia Institute of Technology. 1999

[15] Bowman, D., D. Johnson, and L. Hodges. *Testbed Evaluation of VE Interaction Techniques*. in *Proceedings of ACM VRST*. 1999.

[16] Bowman, D., D. Koller, and L. Hodges, *Evaluation of Movement Control Techniques for Immersive Virtual Environments.* 1997.

[17] Bowman, D., D. Koller, and L. Hodges. *Travel in Immersive Virtual Environments: An Evaluation of Viewpoint Motion Control Techniques*. in *Virtual Reality Annual International Symposium*. 1997: IEEE CS Press, Los Alamitos, Calif.

[18] Bowman, D., D. Koller, and L. Hodges, *A Methodology for the Evaluation of Travel Techniques for Immersive Virtual Environments.* Virtual Reality: Research, Development and Applications, 1998. **3**(2): p. 120-131.

[19] Bowman, D., E. Kruijff, and J. LaViola, *An Introduction to 3D User Interface Design.* Presence: Teleoperators and Virtual Environments, 2001. **10**(1): p. 96-108.

[20] Brooks, F.P. *Grasping Reality Through Illusion - Interactive Graphics Serving Science*. in *ACM CHI 1988 Conference on Human Factors in Computing Systems*. 1988: ACM.

[21] Bruemmer, D.J., J.L. Marble, D.D. Dudenhoeffer, M.O. Anderson, and M.D. McKay. *Mixed-Initiative Control for Remote Characterization of Hazardous Environments*. in *HICSS*. 2003. Waikoloa Village, HI.

[22] Brusilovsky, P., *Efficient techniques for adaptive hypermedia*, in *Intelligent Hypertext: Advanced Techniques for the World Wide Web*, C. Nicholas and J. Mayfield, Editors. 1997, Springer-Verlag: Berlin. p. 12-30.

[23] Burke, R., *Hybrid Recommender Systems: Survey and Experiments.* User Modeling and User-Adapted Interaction, 2002. **12**(4): p. 331-370.

[24] Burtnyk, N., A. Khan, G. Fitzmaurice, R. Balakrishnan, and G. Kurtenback. *StyleCam: Interactive Stylized 3D Navigation Using Integrated Spatial and Temporal Controls*. in *Symposium on User Interface Software and Technology*. 2002. Paris, France.

[25] Card, S.K., J.D. Mackinlay, and B. Shneiderman, *Readings in Information Visualization: Using Vision to Think*. 1999, San Francisco, Calif.: Morgan Kaufmann Publishers.

[26] Chen, M., S.J. Mountford, and A. Sellen, *A study in interactive 3D Rotation using 2D Control Devices.* Computer Graphics, 1988. **22**(4): p. 121-130.

[27] Chen, S.E. *QuickTime VR - An Image Based Approach to Virtual Environment Navigation*. in *Proceedings of ACM SIGGRAPH*. 1995: ACM.

[28] Cheng, P., R. Lowe, and M. Scaife, *Cognitive Science Approaches to Understanding Diagrammatic Representations.* Artificial Intelligence Review, 2001. **15**: p. 79-94.

[29] Chi, E.H., *A Framework for Visualizing Information*. Human-Computer Interaction Series. 2002, Netherlands: Kluwer Academic Publishers.

[30] Chittaro, L. and P. Coppola. *Animated Products as a Navigation Aid for E-commerce*. in *Proceedings of CHI2000: ACM Conference on Human Factors in Computing Systems*. 2000. The Hague, The Netherlands: ACM Press.

[31] Chittaro, L., L. Ieronutti, and R. Ranon, *Navigating 3D Virtual Environments by Following Embodied Agents: A Proposal and Its Informal Evaluation on a Virtual Museum Application.* PsychNology Journal (Special Issue on Human-Computer Interaction), 2004. **2**(1): p. 24-42.

[32] Cleveland, W.S., *The Elements of Graphing Data*. 1985, Montery, CA: Wadsworth Advanced Books and Software.

[33] Colle, H. and G. Reid, *The Room Effect: Metric Spatial Knowledge of Local and Separated Regions.* Presence: Teleoperators and Virtual Environments, 1998. **7**(2): p. 116-128.

[34] Cruz-Neira, C., D.J. Sandin, and T. DeFanti, *Surround-Screen Projection-Based Virtual Reality: The Design and Implementation of the CAVE.* ACM Computer Graphics, 1993. **27**(2): p. 135-142.

[35] Cutting, J.E., *How the Eye Measures Reality and Virtual Reality*. Behavior Research Methods, Instruments & Computers, 1997. **29**(1): p. 27-36.

[36] Czerwinski, M., D.S. Tan, and G.G. Robertson. *Women Take a Wider View*. in *ACM CHI 2002 Conference on Human Factors in Computing Systems*. 2002.

[37] Darken, R., K. Kempster, and B. Peterson. *Effects of Streaming Video Quality of Service on Spatial Comprehension in a Reconnaissance Task.* in *Proceedings of the meeting of I/ITSEC*. 2001.

[38] Darken, R. and B. Peterson, *Spatial Orientation, Wayfinding and Representation.*, in *Handbook of Virtual Environment Technology*, K. Stanney, Editor. 2001, Lawrence Erlbaum Associates: Mahway, NJ.

[39] Darken, R. and J.L. Siebert, *Navigating Large Virtual Spaces.* International Journal of Human-Computer Interaction, 1996. **8**(1): p. 49-72.

[40] Darken, R. and J.L. Siebert. *Wayfinding Strategies and Behaviors in Large Virtual Worlds*. in *ACM CHI 1996 Conference on Human Factors in Computing Systems*. 1996.

[41] Datey, A. Experiments in the Use of Immersion for Information Visualization. Masters. Computer Science, Viginia Polytechnic Unversity. 2002

[42] Dennis, B.M. and C. Healey, *Assisted Navigation for Large Information Spaces.* 2002.

[43] Draper, M.H., E.S. Viirre, T.A. Furness, and V.J. Gawron, *Effects of Image Scale and System Time Delay on Simulator Sickness with Head-Coupled Virtual Environments.* Human Factors, 2001. **43**(1): p. 129-146.

[44] Drascic, D. and P. Milgram. *Perceptual Issues in Augmented Reality*. in *Proceedings of the Stereoscopic Displays and Virtual Reality Systems III*. 1996: SPIE.

[45] Drucker, S.M. and D. Zeltzer. *Intelligent Camera Control in a Virtual Environment*. in *Proceedings of Graphics Interface '94*. 1994: Canadian Information Processing Society, Banff, Alberta, Canada.

[46] Drury, J., J. Scholtz, and H. Yanco. *Awareness in Human-Robot Interactions*. in *IEEE Conference on Systems, Man and Cybernetics*. 2003. Washington, DC.

[47] Dunlop, R., Introduction to Catmull-Rom Splines. http://www.mvps.org/directx/articles/catmull/ accessed: 2004

[48] Endsley, M.R., *Toward a Theory of Situation Awareness in Dynamic Systems.* Human Factors, 1995. **37**(1): p. 32-64.

[49] Fong, T. and C. Thorpe, *Vehicle Teleoperation Interfaces.* Autonomous Robots, 2001(11): p. 9-18.

[50] Fong, T., C. Thorpe, and C. Baur, *Advanced Interfaces for Vehicle Teleoperation: Collaborative Control, Sensor Fusion Displays, and Remote Driving Tools.* Autonomous Robots, 2001(11): p. 77-85.

[51] Fong, T., C. Thorpe, and C. Baur, *Robot, Asker of Questions.* Robotics and Autonomous Systems, 2003(42): p. 235-243.

[52] Gabbard, J. and D. Hix. *A Taxonomy of Usability Characteristics in Virtual Environments* 1997: ONR

[53] Galyean, T. *Guided Navigation of Virtual Environments*. in *1995 Symposium on Interactive 3D Graphics*. 1995.

[54] Goerger, S., R. Darken, M. Boyd, T. Gagnon, S. Liles, J. Sullivan, and J. Lawson. *Spatial Knowledge Acquisition from Maps and Virtual Environments in Complex Architectural Spaces.* in *16th Applied Behavioral Sciences Symposium*. 1998. U.S. Air Force Academy, Colorado Springs, CO.

[55] Golledge, R.G., *Human Wayfinding and Cognitive Maps*, in *Wayfinding Behavior. Cognitive Mapping and Other Spatial Processes*, R.G. Golledge, Editor. 1999, Johns Hopkins University Press: Baltimore, MD. p. 5-45.

[56] Haber, R.B. and D.A. McNabb, *Visualization Idioms: A Conceptual Model for Scientific Visualization Systems*, in *Visualization in Scientific Computing*, B. Shriver, G.M. Nielson, and L.J. Rosenblum, Editors. 1990, IEEE Computer Society Press.

[57] Halper, N. and P. Oliver. *CamPlan: A Camera Planning Agent*. in *AAAI Workshop on Smart Graphics*. 2000. Stanford: AAAI Press, Menlo Park.

[58] Hamming, R.W., *Numerical Methods for Scientists and Engineers.* 1962, New York: McGraw-Hill.

[59] Hanson, A. and E. Wernert. *Constrained 3D Navigation with 2D controllers*. in *Visualization '97*. 1997: IEEE Computer Society Press, Los Alamitos, CA.

[60] Hanson, A., E. Wernert, and S. Hughes, *Constrained Navigation Environments*, in *Scientific Visualization: Dagstuhl '97 Proceedings*, H. Hagen, G.M. Nielson, and F. Post, Editors. 1999, IEEE Computer Society Press. p. 95-104.

[61] He, L.w., M.F. Cohen, and D.H. Salesin. *The virtual cinematographer: A paradigm for automatic real-time camera control and directing*. in *SIGGRAPH 96*. 1996: ACM Press, New York NY.

[62] Hearst, M.A., J. Allen, E. Horvitz, and C. Guinn, *Trends & Controversies: Mixed-initiative interaction.* IEEE Intelligent Systems, 1999. **14**(4): p. 14-23.

[63] Hinkley, K., R. Pausch, J. Goble, and N. Kassell. *A survey of design issues in spatial input*. in *Proceedings of ACM Symposium on User Interface Software and Technology (UIST)*. 1994. Marina Del Rey, CA.

[64] Hirtle, S. and J. Hudson, *Acquisition of Spatial Knowledge for Routes.* Journal of Environmental Psychology, 1991. **11**: p. 335-345.

[65] Hix, D., E. Swan, J. Gabbard, M. McGee, J. Durbin, and T. King. *User-Centered Design and Evaluation of a Real-Time Battlefield Visualization Virtual Environment*. in *IEEE Virtual Reality '99*. 1999: IEEE Computer Society Washington, DC.

[66] Hübscher, R. and S. Puntambekar. *Adaptive Navigation for Learners in Hypermedia is Scaffolded Navigation*. in *Adaptive Hypermedia and Adaptive Web-Based Systems*. 2002. Malaga, Spain: Springer-Verlag.

[67] Hughes, S., P. Brusilovsky, and M. Lewis. *Adaptive Navigation Support in 3D E-commerce Activities*. in *Workshop on Recommendation and Personalization in eCommerce at the 2nd International Conference on Adaptive Hypermedia and Adaptive Web Based Systems*. 2002. Malaga, Spain.

[68] Hughes, S. and M. Lewis. *Attentive Interaction Techniques for Searching Virtual Environments*. in *Human Factors and Ergonomics Society's 46th Annual Meeting*. 2002. Baltimore, MD.

[69] Hughes, S. and M. Lewis. *Directing Attention in Open Scenes*. in *Human Factors and Ergonomics Society's 46th Annual Meeting*. 2002. Baltimore, MD.

[70] Hughes, S. and M. Lewis. *Robotic Camera Control for Remote Exploration*. in *ACM CHI 2004 Conference on Human Factors in Computing Systems*. 2004. Vienna, Austria.

[71] Igarashi, T., R. Kadobayashi, K. Mase, and H. Tanaka. *Path Drawing for 3D Walkthrough*. in *UIST'98*. 1998: ACM Press.

[72] Kay, J. *STRIPE: Remote Driving Using Limited Image Data*. in *CHI: ACM Conference on Human Factors in Computing Systems*. 1995.

[73] Keahey, T.A. *The Generalized Detail-in-Context Problem*. in *IEEE Symposium on Information Visualization*. 1998: IEEE Computer Society Washington, DC.

[74] Keahey, T.A. and E. Roberston. *Nonlinear magnification fields*. in *IEEE Symposium on Information Visualization*. 1997.

[75] Kiss, S. and A. Nijholt. *Viewpoint Adaptation during Navigation based on Stimuli from the Virtual Environment*. in *Proceedings Web3D 2003 Symposium. 8th International Conference on 3D Web Technology*. 2003. Saint Malo,: ACM SIGGRAPH.

[76] Kosslyn, S., *Understandin Charts and Graphs.* Applied Cognitive Psychology, 1989. **3**: p. 185-226.

[77] Kuzuoka, H., K. Yamazaki, A. Yamazaki, K. Jun'ichi, Y. Suga, and C. Heath. *Dual Ecologies of Robot as Communication Media: Thoughts on Coordinating Orientations and Projectability*. in *ACM CHI 2004 Conference on Human Factors in Computing Systems*. 2004. Vienna, Austria.

[78] Larkin, J., ed. *Display-Based Problem Solving*. Complex Information Processing: The Impact of Herbert A. Simon, ed. D. Klahr and K. Kotovsky. 1989, Lawrence Erlbaum Associates: Hillsdale, New Jersey. 319-341.

[79] Larkin, J. and H.A. Simon, *Why a diagram is (sometimes) worth 10,000 words.* Cognitive Science, 1987. **11**: p. 65-99.

[80] Lee, J. and N. Moray, *Trust, Control Strategies, and Allocation of Function in Human-Machine Systems.* Ergonomics, 1992. **35**(10): p. 1243-1270.

[81] Loomis, J., R. Klatzky, R.G. Golledge, and J.W. Philbeck, *Human Navigation by Path Integration*, in *Wayfinding Behavior. Cognitive Mapping and Other Spatial Processes*, R.G. Golledge, Editor. 1999, Johns Hopkins University Press: Baltimore, MD. p. 125-151.

[82] Mackinlay, J., S. Card, and G. Roberston, *Rapid Controlled Movement through a Virtual 3D Workspace.* Computer Graphics, 1990. **24**(4): p. 171-166.

[83] Masliah, M. and P. Milgram. *Measuring the allocation of control in a 6 degree-of-freedom docking experiment*. in *Conference on Human Factors in Computing Systems*. 2000. The Hague, The Netherlands.

[84] McAllister, D.F., *3D Displays*, in *Wiley Encyclopedia on Imaging*. 2002. p. 1327-1344.

[85] McCormick, B., T. DeFanti, and M. Brown, *Visualization in Scientific Computing - a synopsis.* IEEE Computer Application Graphics, 1987. **7**: p. 61-70.

[86] McGovern, D.E. *Experiences and Results in Teleoperation of Land Vehicles*.SAND 90-0299. 1990. Albuquerque, NM: Sandia National Laboratories

[87] Merriam-Webster, *Collegiate Dictionary (10th ed)*. 1993, Springfield, MA: Merriam-Webster.

[88] Merrill, M.D., L. Zhongmin, and M.K. Jones, *Instructional Transaction Shells: Responsibilities, Methods, and Parameters.* Instructional Technology, 1992(February): p. 5-26.

[89] Milgram, P. and H. Colquhoun, *A Taxonomy of Real and Virtual World Display Integration*, in *Mixed Reality - Merging Real and Virtual Worlds.*, Y.O.a.H. Tamura, Editor. 1999, Springer Verlag: Berlin. p. 1-16.

[90] Mine, M. *Virtual Environment Interaction Techniques* 1995: UNC Chapel Hill Computer Science Technical Report TR95-018

[91] Muir, B., *Trust Between Humans and Machines, and the Design of Decision Aids.* International Journal of Man-Machine Studies, 1987. **27**(5-6): p. 527-539.

[92] Munzner, T. Interactive Visualization of Large Graphs and Networks. Computer Science, Stanford University. 2000

[93] Murphy, R., Robot-Assisted Search and Rescue at the WTC Disaster. www.csee.usf.edu/~murphy/NSF-DC.ppt accessed: March 2005

[94] Neale, D.C. and J.M. Carroll, *The Role of Metaphors in User Interface Design*, in *Handbook of Human Computer Interaction*, M.G. Helander, T.K. Landauer, and P. Prabhu, Editors. 1997, Elsevier Science.

[95] Nieuwenhuisen, D. and M.H. Overmars. *Motion Planning for Camera Movements in Virtual Environments* 2002. Utrecht, the Netherlands: Utrecht University Information and Computing Sciences

[96] Ohmi, M. *Roles of Additional Information for Wayfinding in Virtual Environments.* in *Eighth International Conference on Artificial Reality and Tele-Existence*. 1999.

[97] OpenGL, www.opengl.org. accessed: 2004

[98] Ou, S., D.R. Karuppiah, A.H. Fagg, and E. Riseman. *An Augmented Virtual Reality Interface for Assistive Monitoring of Smart Spaces*. in *IEEE International Conference on Pervasive Computing and Communications*. 2004.

[99] Parker, G., G. Franck, and C. Ware, *Visualization of Large Nested Graphs in 3D: Navigation and Interaction.* Journal of Visual Languages and Computing, 1998. **9**(3): p. 299-317.

[100] Peruch, P., J. Vercher, and G. Guthier, *Acquisition of Spatial Knowledge through Visual Exploration of Simulated Environments.* Ecological Psychology, 1995. **7**(1): p. 1-20.

[101] Plumlee, M. and C. Ware. *An Evaluation of Methods for Linking 3D Views*. in *Symposium on Interactive 3D Graphics*. 2003.

[102] Robertson, G.G., *Information Visualization using 3D Interactive Animation.* Communications of the Association for Computing Machinery (CACM), 1993. **26**(4): p. 57-71.

[103] Robertson, G.G., J.D. Mackinlay, and S.K. Card. *Cone Trees: Animated 3D Visualizations of Hierarchical Information*. in *ACM SIGCHI Conference on Human Factors in Computing Systems*. 1991: ACM Press.

[104] Roth, S. and J. Mattis. *Data Characterization for Intelligent Graphics Presentation*. in *CHI '90*. 1990.

[105] Satalich, G. Navigation and Wayfinding in Virtual Reality: Finding Proper Tools and Cues to Enhance Navigation Awareness. Master's. University of Washington. 1995

[106] Scaife, M. and Y. Rogers, *External Cognition: How do Graphical Representation Work?* International Journal of Human-Computer Studies, 1996. **45**: p. 185-213.

[107] SGI, The Cosmo Player. http://www.sgi.com/software/cosmo/player.html. accessed:

[108] Shepard, R.N. and J. Metzler, *Mental Rotation of Three-Dimensional Objects.* Science, 1971. **171**: p. 702-303.

[109] Shneiderman, B., *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. 1993: Addison -Wesley.

[110] Shneiderman, B., *Dynamic Queries for Visual Information Seeking.* IEEE Software, 1994. **6**(11): p. 70-77.

[111] Shneiderman, B. *The eyes have it: A task by data type taxonomy for information visualizations*. in *IEEE Visual Languages*. 1996. Boulder, CO.

[112] Siegel, A.W. and S. White, *The development of spatial representations in large-scale environments*, in *Advances in child development and behavior,* H.W. Reese, Editor. 1975, New York: Academic Press. p. 9-55.

[113] Slater, M. and M. Usoh. *Presence in Immersive Virtual Environments*. in *IEEE VRAIS*. 1993.

[114] Smelser, N.J. and P.B. Baltes, eds. *International Encyclopedia of the Social and Behavioral Sciences*. 2001, Pergamon Press: Oxford.

[115] Spence, R., *Information visualization*. 2001, Harlow: Addison-Wesley. xvii, 206.

[116] Spring, M., *Personal Communication*. 2003.

[117] Stanney, K., R. Mourant, and R. Kennedy, *Human Factors Issues in Virtual Environments: A review of the literature.* Presence, 1998. **7**(4): p. 327-351.

[118] Stevenson, B., ed. *The Home Book of Proverbs, Maxims and Familiar Phrases.* 1948, MacMillan: New York.

[119] Stoakley, R., M. Conway, and R. Pausch. *Virtual Reality on a WIM: Interactive Worlds in Miniature*. in *CHI: ACM Conference on Human Factors in Computing Systems*. 1995.

[120] Tan, D.S., G.G. Robertson, and M. Czerwinski. *Exploring 3D Navigation: Combining Speed-coupled flying with orbiting*. in *CHI 2001 Conference on Human Factors in Computing Systems*. 2001. Seattle, WA.

[121] Thomas, G., C. Petrie, N. Loeppke, M. Bauerly, R. Mills, M. Rick, B. Wyatt, R. Reagan, N. Cabrol, S. Dow, S. Fischer, S. McClarigan, J. Steele, and J. Wagner. *Project MARVIN: Mars Advance Robotic Visualization Initiative*. in *Iowa Space Grant Consortium Conference*. 1999.

[122] Thorndyke, P.W. and B. Hayes-Roth, *Differences in spatial knowledge acquired from maps and navigation.* Cognitive Psychology, 1982. **14**: p. 560-589.

[123] Tufte, E.R., *The Visual Display of Quantitative Information*. 1983: Graphics Press.

[124] Tversky, B., *Cognitive Maps, Cognitive Collages and Spatial Mental Models*, in *Spatial Information Theory: Theoretical basis for GIS*, A.U. Frank and I. Campari, Editors. 1993, Springer-Verlag: Heidelberg-Berlin.

[125] van Dijk, B., R. op den Akker, A. Nijholt, and J. Zwiers, *Navigation Assistance in Virtual Worlds*, in *Informing Science, Special Series on Community Informatics*, E. Rathswohl and C. Winer, Editors. 2003. p. 115-125.

[126] Varian, H.R., *The Information Economy.* Scientific American, 1995. **273**(3): p. 200-201.

[127] Viega, J., M. Conway, G. Williams, and R. Pausch. *3D Magic Lenses*. in *Symposium on User Interface Software and Technology*. 1996. Seattle, WA.

[128] Vygotsky, L.S., *Mind in Society*. 1978, Cambridge: Harvard University Press.

[129] Wang, J., M. Lewis, S. Hughes, M. Koes, and S. Carpin. *Validating USARsim for use in HRI Research*. in *Human Factors and Ergonomics Society's 49th Annual Meeting*. 2005. Orlando, FL.

[130] Ware, C., *Using Hand Position for Virtual Object Placement.* Visual Computer, 1990(6): p. 245-253.

[131] Ware, C., *Information visualization : perception for design*. 2000, San Francisco: Morgan Kaufman. xxiii 438.

[132] Ware, C. and S. Osborne. *Exploration and Virtual Camera Control in Three Dimensional Environments*. in *Proceedings of the 1990 Symposium on Interactive 3D Graphics*. 1990: ACM Press New York, NY.

[133] Web3D_Consortium, The Virtual Reality Modeling Language. http://www.web3d.org/x3d/specifications/vrml/ISO_IEC_14772-All/ accessed: 2004

[134] Web3D_Consortium, Extensible 3D (X3D): ISO/IEC 19775. http://www.web3d.org/x3d/specifications/ISO-IEC-19775/ accessed: 2004

[135] Wehrend, S. and C. Lewis. *A Problem-Oriented Classification of Visualization Techniques*. in *Visualization '90*. 1990.

[136] Werner, S., B. Krieg-Bruckner, H. Mallot, K. Schweizer, and C. Freksa, *The Role of Landmark, Route, and Survey Knowledge in Human and Robot Navigation*, in *Informatik '97*, M. Jarke, K. Pasedach, and K. Pohl, Editors. 1997, Springer: Berlin. p. 41-50.

[137] Wernert, E. and A. Hanson. *A framework for assisted exploration with collaboration*. in *Visualization '99*. 1999: IEEE Computer Society Press.

[138] Wickens, C.D., J.D. Lee, Y. Liu, and S. Becker, *An Introduction to Human Factors Engineering*. 2nd ed. 2004, Upper Saddle River, NJ: Pearson, Prentice Hall.

[139] Witmer, B.G. and M.J. Singer, *Measuring Presence in Virtual Environments: A Presence Questionnaire.* Presence: Teleoperators and Virtual Environments, 1998. **7**(3): p. 225-240.

[140] Woods, D.D., J.S. Tittle, M. Feil, and A. Roesler, *Envisioning Human-Robot Coordination in Future Operations.* Systems, Man and Cybernetics, Part C: Applications and Reviews, 2004. **34**(2): p. 210-218.

[141] Yanco, H., J. Drury, and J. Scholtz, *Beyond Usability Evaluation: Analysis of Human-Robot Interaction.* Journal of Human-Computer Interaction, 2004. **19**: p. 117-149.

[142] Yeh, M., C.D. Wickens, and F.J. Seagull. *Effects of Frame of Reference and Viewing Condition on Attentional Issues with Helmet Mounted Displays*.ARL-98-1. 1998. Urbana-Champaign: University of Illinois at Urbana-Champaign

[143] Zhai, S., *User Performance in Relation to 3D Input Device Design.* Computer Graphics, 1998. **32**(4): p. 50-54.

[144] Zhai, S., J. Wright, T. Selker, and S. Kelin. *Graphical Means of Directing Users' Attention in the Visual Interface*. in *Interact '97*. 1997: Chapman & Hall, Ltd, London.

[145] Zhou, M.X. and S. Feiner. *Visual Task Characterization for Automated Visual Discourse Synthesis*. in *ACM CHI*. 1998. Los Angeles, CA.